

2015

Engineering yeast genomes and populations

<https://hdl.handle.net/2144/13712>

Boston University

BOSTON UNIVERSITY
COLLEGE OF ENGINEERING

Dissertation

ENGINEERING YEAST GENOMES AND POPULATIONS

by

JAMES EDWARD DICARLO

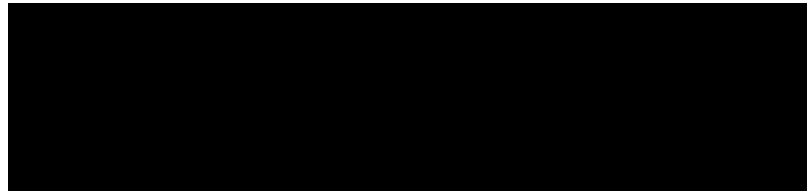
B.S., Johns Hopkins University, 2010
M.S., Boston University, 2014

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2015

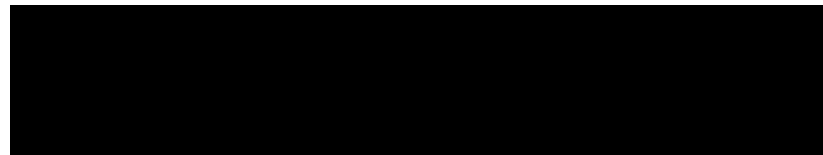
Approved by

First Reader



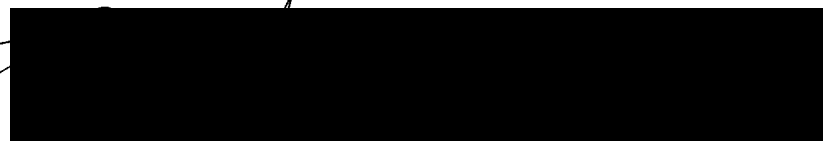
George Church, Ph.D.
Professor of Genetics
Harvard University, School of Medicine

Second Reader



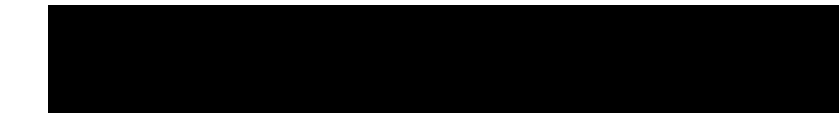
Ahmad S. Khalil, Ph.D.
Assistant Professor of Biomedical Engineering

Third Reader



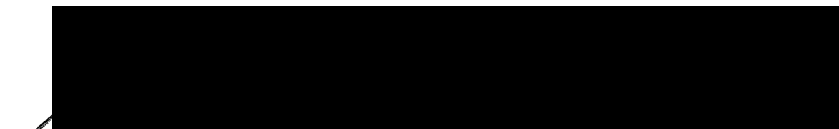
James J. Collins, Ph.D.
Termeer Professor of Bioengineering
Massachusetts Institute of Technology

Fourth Reader



James E. Galagan, Ph.D.
Associate Professor of Biomedical Engineering
Associate Professor of Microbiology

Fifth Reader



Wilson Wong, Ph.D.
Assistant Professor of Biomedical Engineering

ACKNOWLEDGMENTS

First of all, I want to thank my advisor George Church for being a constant source of inspirational and pragmatic advice for my both my experiments and my future career as a scientist. In his lab I was allowed an incredible amount of creative freedom. In the Church lab, thinking outside of the box is expected and the environment supports this both scientifically and socially/emotionally. Working with George has been a privilege and I have grown immensely as a scientist under his mentorship. I will always be grateful for the opportunity to work with George. I also want to thank Harris Wang for being my first mentor in the Church lab. Harris sent me off in the right direction for my PhD and I learned an incredible amount about smart experimental design from working with him.

I am thankful to all the members of the Church lab for providing stimulating scientific conversations as well as a ton of fun. I want to especially thank Xavier Rios, Jon Lim, Nikolai Eroshenko, Alice McElhinney, and Dan Mandell for going above and beyond as friends, listening to my terrible jokes, and always being there lean on. For all the support they have provided over the years I also want to thank Raj Chari, Michael Napolitano, Mike Mee, Marc Lajoie, Sandrine Boissel, George Chao, Justin Feng, Henry Lee, Jaimie Rogers, Jonathan Scheimann, Peter Smith, Chuck Nobleston, Noah Taylor, Arthur Yu, Kevin Esvelt, Sven Dietz, Chris Gregg, Nick Guido, Graham Rockwell, Sri Kosuri, Uri Laserson, Madeline Price-Ball, Poyi Huang, Josh Mosberg, Sara Vassallo, Margo Monroe, Adrian Briggs, Julie Norville, Matthieu Landon, Alejandro Chavez, Susan Bryne, Prashant Mali, Jun Li, Wei-Leong Chew, Laura Glass, Yveta Masar, and

John Aach. Also, my close friend Dan Bisers has supported me through thick-and-thin and I would not have finished or gone to graduate school without his emotional support.

My journey to become a scientist would not have been possible without an extended web of incredibly supportive family, friends, teachers and mentors. Most importantly, I would like to thank my parents, Valerie Lithotomos and Edward DiCarlo who have always encouraged my scientific interests, hobbies and pets. Whether it was a helping me maintain too many aquariums, or supporting the many reptiles and mammals that I shared my youth with, my parents were always a source of loving encouragement. I never felt unloved. I always know that they will be there for me, as I will be there for them. My parents also made sure I had an amazing education and of the educational opportunities they afforded me, I developed a love for academic science and was able to meet some of my most important science mentors. I have been incredibly lucky to have a supportive and loving extended family as well.

My Aunts, Uncles, and grandparents have always encouraged my love of science have been huge emotional as well as comedic supports throughout my life. My Aunt Cheryl DiCarlo and Uncle Patrick Roache were always huge supporters for my career in science and for that I am very grateful. I have always looked up to my cousin Laura Hack, who continues to be my gold standard as a scientist, physician and as a person. My high school biology teachers Mrs. Cradler and Mrs. Touhey inspired and supported my enthusiasm for biology, even though it may have been too much for a busy teacher to handle at the time. They helped guide me towards my love of molecular biology and sent me into college with a well-formed understanding of the area my future career. In

college, Jef Boeke let me work in his lab and was one of my most influential scientific mentors. He and Lisa Scheifele introduced me to the world of synthetic biology and yeast. I learned how to use a pipette skillfully and think critically about experiments in the Boeke lab.

ENGINEERING YEAST GENOMES AND POPULATIONS

JAMES EDWARD DICARLO

Boston University College of Engineering, 2015

Major Professor: George Church, Ph.D., Professor of Genetics, Harvard University

ABSTRACT

The field of synthetic biology seeks to use design principles of life to create new genes, organisms and populations to both better understand biology as well as generate species with useful properties. Budding yeast has been a workhorse for synthetic biology, as well as an important model organism in the broader fields of molecular biology and genetics. This thesis aimed to create genome engineering tools for the manipulation of genomes, with direct applications in yeast. I focused developing high-throughput and highly efficient methods for making genomic modifications in yeast to allow for the generation of large libraries of precisely modified yeast genomes.

By manipulation of endogenous DNA recombinases and mismatch repair enzymes in yeast, we were able to develop an oligonucleotide only method for genome engineering to generate libraries as large as 10^5 individuals with a frequency of modification as high as 1%. Additionally, we validated the use of RNA-guided CRISPR/Cas9 endonucleases to make changes in yeast genomes, resulting in frequencies of genome modification >90% in transformed populations. We further optimized this method to generate larger libraries as high as 10^5 individuals and explored a proof of concept epistasis experiment involving thermotolerance. Lastly, the propagation of changes to successive generations is useful when engineering organisms on the

population level. To this end we explored the use of RNA-guided gene drives to bias inheritance in *S. cerevisiae*. We show that inheritance of these selfish elements can be biased to over 99% and is reversible.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
ABSTRACT.....	vii
TABLE OF CONTENTS.....	ix
LIST OF TABLES	xiii
LIST OF FIGURES	xiv
LIST OF ABBREVIATIONS.....	xvi
CITATIONS TO PREVIOUSLY PUBLISHED WORK.....	xvii
CHAPTER 1 - Introduction	1
1.1 Motivation.....	1
1.2 High Efficiency Oligonucleotide-mediated Genome Modification.....	4
1.3 Using CRISPR-Cas9 for Genome Engineering	4
1.4 RNA-guided Gene Drives.....	5
CHAPTER 2 – Yeast Oligonucleotide-mediated Engineering	7
2.1 Introduction.....	7
2.2 Mismatch Repair Knockout	8
2.3 DNA Recombinase Overexpression	9
2.4 External Factors Affecting Oligonucleotide Recombination.....	12
2.5 Locus Variability	14
2.6 Multiplex Modifications and Cycling.....	15

2.7 Discussion and Future Considerations.....	16
2.8 Figures.....	20
CHAPTER 3 – Genome engineering in yeast using CRISPR-Cas9.....	25
3.1 Introduction.....	25
3.2 gRNA Expression and Genomic Target Identification.....	26
3.3 CRISPR-Cas directed CAN1 mutagenesis	27
3.4 CRISPR-Cas stimulated homologous recombination with donor DNA and transient gRNA PCR cassette	29
3.5 CRISPR-Cas stimulated homologous recombination with donor DNA and gRNA expression plasmid.....	30
3.6 Materials and Methods.....	32
3.6.1 Strains and Media.	32
3.6.2 Plasmid Construction.	32
3.6.3 Transformation of plasmids for CAN1 mutagenesis.	33
3.6.4 Electroporation of Transient gRNA cassette and oligonucleotides.	34
3.6.5 Transformation of gRNA plasmid with donor DNA into Cas9 expressing cells.	35
3.6.6 Galactose induction of Cas9.	35
3.6.7 Toxicity of CRISPR system.....	36
3.6.8 Identification of CRISPR-Cas targets in yeast genome.	36
3.7 Discussion and Future Consideration	37
3.8 Figures.....	40

CHAPTER 4 – Optimization and Application of Cas9-mediated genome engineering...	43
4.1 Introduction.....	43
4.2 Multiplex Oligonucleotide Genome Editing via Cas9 with cycling.....	44
4.3 Optimization of CRISPR-Cas9 for oligonucleotide replacement and Donor DNA Delivery.....	45
4.4 Strain Engineering of Industrially Relevant Phenotypes	46
4.4.1 Thermotolerance	46
4.5 Figures.....	48
Chapter 5 – RNA-guided gene drives can efficiently and reversibly bias inheritance in <i>S. cerevisiae</i>	51
5.1 Introduction.....	51
5.2 Molecularly Contained Gene Drives	52
5.3 Gene Drives Spread into Phylogenetically Diverse Yeast Strains.....	55
5.4 Complete Gene Drives and Reversibility	56
5.5 Discussion and Impact on the Gene Drive Community.....	57
5.6 Materials and Methods.....	60
5.6.1 Physical Confinement of Gene Drives.....	60
5.6.2 Plasmids and genomic cassettes.....	61
5.6.3 Sporulation and tetrad dissection	63
5.6.4 Selection for URA3 function	63
5.6.5 Verification of chromosomal segregation.....	63
5.6.6 Quantitative PCR	64

5.6.7 qPCR Calculations	64
5.7 Figures.....	66
CHAPTER 6 - Discussion	71
APPENDIX.....	73
BIBLIOGRAPHY	74

LIST OF TABLES

Table A1. List of matings for yeast gene drive experiments.. ..	73
---	----

LIST OF FIGURES

Figure 2.1 Diagram of Yeast Oligo-mediated Genome Engineering (YOGÉ) applications	20
Figure 2.2 Oligonucleotide recombination frequency across all strains and loci tested...	21
Figure 2.3 External factors affecting oligonucleotide recombination.	22
Figure 2.4 Variability in oligonucleotide recombination across DNA strands, loci and strains	23
Figure 2.5 Cycling and multiplex modifications using YOGÉ in strain CEN.PK.RR at the CAN1 and LYP1 loci.....	24
Figure 3.1 Diagram of Cas9 complex and schematic of genetic constructs.	40
Figure 3.2 Figure 3.2 CRISPR mediated genomic mutagenesis.....	41
Figure 3.3 Transient Transformation of gRNA cassette and oligonucleotide recombination in Cas9 constitutively expressing strains	42
Figure 4.1 Optimization of Multiplex Genome engineering protocol	48
Figure 4.2 Optimization of gRNA/Cas9 plasmid and donor double stranded oligonucleotide delivery.....	49
Figure 4.3 Thermotolerant mutant phenotypes and multiplex cassette insertion	50
Figure 5.1 Mechanism and population-level effect of endonuclease gene drive.....	66
Figure 5.2 Biased inheritance of ADE2 is readily visible in <i>S. cerevisiae</i>	67
Figure 5.3 Gene drives and cargo genes remain intact upon copying and can spread by targeting both non-essential and essential genes	68
Figure 5.4 qPCR data for yeast gene drives.....	69

Figure 5.5 Essential, autonomous, and reversal gene drives	70
--	----

LIST OF ABBREVIATIONS

COD	Coding
CRISPR.....	Clustered Regularly Interspaced Short Palindromic Repeats
DNA.....	Deoxyribonucleic Acid
DSB.....	Double-strand break
gRNA	Guide RNA
HR.....	Homologous recombination
MMR.....	Mismatch Repair
NCOD	Noncoding
NHEJ.....	Non-homologous End Joining
RNA	Ribonucleic Acid
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SC.....	Synthetic Complete
YOGF	Yeast Oligo-mediated Genome Engineering
YPAD.....	Yeast Peptone Adenine Dextrose
YPD.....	Yeast Peptone Dextrose

CITATIONS TO PREVIOUSLY PUBLISHED WORK

The majority of the text in this dissertation is reproduced from published material.

Changes made in this work were done to reflect the theme of the dissertation.

Chapter 2 was published as:

DiCarlo, James E., Andrew J. Conley, Merja Penttilä, Jussi Jäntti, Harris H. Wang, and George M. Church, ‘Yeast Oligo-Mediated Genome Engineering (YOGE)’, *ACS Synthetic Biology*, 2 (2013), 741–49 <<http://dx.doi.org/10.1021/sb400117c>>

Chapter 3 was published as:

DiCarlo, James E., Julie E. Norville, Prashant Mali, Xavier Rios, John Aach, and George M. Church, ‘Genome Engineering in *Saccharomyces cerevisiae* Using CRISPR-Cas Systems’, *Nucleic Acids Research*, 41 (2013), 4336–43
<<http://dx.doi.org/10.1093/nar/gkt135>>

Chapter 5 was published as:

DiCarlo, James E., Alejandro Chavez, Sven L. Dietz, Kevin M. Esvelt, and George M. Church, ‘RNA-guided Gene Drives Can Efficiently and Reversibly Bias Inheritance in Wild Yeast’, *bioRxiv*, 2015, 013896 <<http://dx.doi.org/10.1101/013896>>

CHAPTER 1 - Introduction

1.1 Motivation

Since the discovery of restriction endonucleases in the 1970's scientists have been continuously improving their ability to make precise and efficient manipulations of DNA molecules. (1–3) Modification of DNA in living cells allows scientists and engineers to better understand the design rules that govern biology. (3) The field of synthetic biology seeks to utilize the design principles of organisms to create new genes, organisms and populations to both better understand biology as well as generate organisms with useful properties. (4–6) Many organisms have been the focus of synthetic biology from, *Escherichia coli*, to budding yeast, to mammalian cells. Budding yeast, or *Saccharomyces cerevisiae*, has been an economically important organism since the beginning of organized civilization. The use of yeasts for consumable fermentation products, such as beer and wine, as well as for bread leavening, dates back to near Neolithic times. (7, 8) As a model of the eukaryotic cell, baker's yeast (*Saccharomyces cerevisiae*) was first used in the early 1900's and became a crucial organism in the study of chromosome dynamics and genetic manipulation. (9–11) The first transformation of yeast with extracellular DNA was in 1978 by Hinnen et al., making it one of the first organisms to be genetically modified. Since 1980's, *S. cerevisiae* has had a successful history of genome manipulation due to its efficient endogenous homologous recombination machinery. (12)

In yeast, high frequency genomic recombination using double stranded DNA (approaching 10^{-3} transformants per surviving cell) requires large homology arms with a

minimum length of ~500 base pairs.(13, 14) The use of shorter homology arms (35 – 70 base pairs) drastically reduces recombination frequency to about 10⁻⁶ transformants per surviving cell. (14–16) Directed homologous recombination using single-stranded oligonucleotides has thus far been limited by low recombination frequency in the absence of artificial selection for the modified allele. Nonetheless, short oligonucleotides have significant advantages over large double-stranded DNA constructs since oligonucleotides are inexpensive and easily obtained from commercial sources without the need for PCR amplification and purification.(16, 17) An efficient method for oligonucleotide-mediated recombination engineering (recombineering) in yeast would thus have a variety of applications. Furthermore, high efficiency oligo-mediated recombineering in yeast will further its emerging role as a biological chassis for the de novo assembly and construction of genes, gene clusters, and genomes. (18, 19)

The generation of genetic diversity in a rational, site-specific way would allow for higher fitness mutant libraries that could be screened for desirable phenotypes. Furthermore, easy generation of multiple alleles would allow for investigation of epistatic interactions between DNA regulatory regions and proteins. For example, pathways for desirable compounds could be optimized via promoter tuning using oligonucleotides. (20, 21) Additionally, changes could be made to protein coding regions to generate mutant protein alleles. In addition to oligonucleotides, the genomic insertion of larger DNA with high frequency and precision would allow for easier strain engineering for industrial and academic purposes. With a high frequency integration method, no selection at the site of integration would be required, allowing for higher throughput genome modification,

without a limit of number of selectable markers.

Genomic integration of foreign DNA has many benefits over plasmid introduction. Firstly plasmids, in most cases required selection for maintenance, while genomic integration in many cases is stable and can be maintained continuously without selection. Moreover, expression levels of genomically encoded genes are more constant, and introduction of sequence diversity to alter gene expression at the genomic level generate more reproducible variation. Additionally, genomic manipulation of host genomes allows for perturbation of natural functional genomic elements, which can aid in elucidation of their function.

Lastly, the propagation of changes to successive generations is useful when engineering organisms on the population level. In sexual populations, one method for ensuring the propagation of an allele is by increasing its odds of inheritance. By providing a selective advantage to the allele, that is separate from the gene product, one can ensure it is maintained in a population. One method to bias inheritance in populations is the use of genetic elements called gene drives. (22–24) Frequently, gene drives bias inheritance by encoding an endonuclease that cleaves an endogenous locus to stimulate homologous recombination repair with a locus containing its own coding sequence. Hence upon sexual reproduction, when an uncut locus becomes available, the endonuclease gene drive can repeat the process and spread itself. RNA-guidable endonucleases have allowed for the design of *de novo* endonucleases for any sequence desired in a rapid and simple manner, thus allowing for a gene drive to be designed for any sequence desired. (25) By applying RNA-guided gene drives in modified yeast

alleles are able to be spread in a sexually reproducing population, potentially in a combinatorial manner that would allow for high efficiency and large library design. Because gene drives are able to spread, taking safety precautions is crucial for their study to protect against accidental release.

1.2 High Efficiency Oligonucleotide-mediated Genome Modification

Chapter 1 describes a method to impart genomic changes in *S. cerevisiae* using oligonucleotides with efficiencies ranging from 0.1-1% (meaning 1 in 100 to 1 in 1000 cells in the transformed population have the desired modification). The rate of recombination between the genome and the delivered oligonucleotide was dramatically boosted (by up to 500-fold) via overproduction of endogenous homologous repair proteins and knockout of mismatch repair pathways responsible for removing heterodimeric DNA. This method can generate genomic yeast libraries of the size of 10^5 individuals. (26)

1.3 Using CRISPR-Cas9 for Genome Engineering

Chapter 2 explores the use of CRISPR-Cas9 systems for their applicability in yeast genome engineering. CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) and CRISPR-associated (Cas) systems in bacteria and archaea employ RNA-guided DNA endonuclease activity to provide adaptive immunity against invading foreign nucleic acids. Here we report the use of Type II bacterial CRISPR/Cas system in *Saccharomyces cerevisiae* for genome engineering. The CRISPR components, Cas9 gene and a designer genome targeting CRISPR guide RNA (gRNA), show robust and specific RNA guided endonuclease activity at targeted endogenous loci in yeast. Using a

constitutive Cas9 expression and a transient gRNA cassette we show that targeted double-strand breaks can increase homologous recombination rates of single and double-stranded oligonucleotide donors by 5-fold and 130-fold, respectively. Our approach provides foundations for a simple and powerful genome engineering tool for site-specific mutagenesis and allelic replacement in yeast. (27) Additionally, Chapter 3 discusses an optimized methodology Cas9-mediated genome engineering. We utilized the protocol to explore epistasis in genes involved in heat stress response, and analyzed thermotolerance of generated strains

1.4 Optimization of Cas9-mediated genome engineering

After validation of functionality of CRISPR-Cas9 systems in yeast as well as their potential for genome engineering, we further optimized protocols to use them for genome engineering. We examined plasmid concentrations, as well as double-stranded oligonucleotide concentrations that would that would be yield the highest genome modification frequency. Finally, we completed a proof of concept experiment showing that multiple genomic changes could be made at once and that this process could be cycled. A set of genes involved in thermotolerance was targeted and epistasis between the genes was probed.

1.5 RNA-guided Gene Drives

In Chapter 4, I show molecularly contained gene drive constructs in the yeast *Saccharomyces cerevisiae* that are typically copied at rates above 99% when mated to wild yeast. We successfully targeted both non-essential and essential genes, showed that the inheritance of an unrelated “cargo” gene could be biased by an adjacent drive, and

constructed a drive capable of overwriting and reversing changes made by a previous drive. Our results demonstrate that RNA-guided gene drives are capable of efficiently biasing inheritance when mated to wild-type organisms over successive generations.(28)

CHAPTER 2 – Yeast Oligonucleotide-mediated Engineering

2.1 Introduction

Oligonucleotide-mediated genome engineering is a practical method for performing site-directed mutagenesis to enable the rapid generation of various rationally designed organisms. (21, 29) In *Saccharomyces cerevisiae*, genome manipulation has had a long and successful history due to its efficient endogenous homologous recombination machinery. In general, high frequency genomic recombination using double stranded DNA (approaching 10^{-3} transformants per surviving cell) requires large homology arms with a minimum length of ~500 base pairs. (13, 14) The use of shorter homology arms (35 – 70 base pairs) drastically reduces recombination frequency to about $10^{-6} - 10^{-5}$ transformants per surviving cell. (14–16) Directed homologous recombination using single-stranded oligonucleotides has thus far been limited by low recombination frequency in the absence of artificial selection for the modified allele. Nonetheless, short oligonucleotides have significant advantages over large double-stranded DNA constructs since oligonucleotides are inexpensive and easily obtained from commercial sources without the need for PCR amplification and purification. (16, 17)

An efficient method for oligonucleotide-mediated recombination engineering or allelic replacement in yeast would thus have a variety of applications. Furthermore, high efficiency oligo-mediated recombineering in yeast will further its emerging role as a biological chassis for the *de novo* assembly and construction of genes, gene clusters, and bacterial genomes. (18, 19) Here, we describe yeast strains with highly efficient oligo-mediated recombination capabilities and identify key factors governing these traits for

use in Yeast Oligo-mediated Genome Engineering (YOGЕ) (**Figure 2.1**).

Since homologous recombination rates vary in different yeast strains, we sought to develop YOGЕ in three separate *S. cerevisiae* haploid strains of different lineages (VL6-48, CEN.PK113-7D, and VTT C-68059). VL6-48 is a 288c strain derivative first developed for transformation-associated recombination (TAR) cloning and has been reported to have high transformation efficiency. (30, 31) *S. cerevisiae* CEN.PK113-7D is a lab strain widely used in systems biology and metabolic engineering studies that is also physiologically robust in industrial settings. (32) *S. cerevisiae* (var. *diastaticus*) VTT C-68059 is an industrial yeast strain that was originally isolated as a brewery contaminant. These three strains represent an array of biological backgrounds that may be encountered for general adoption of YOGЕ in different experimental settings.

2.2 Mismatch Repair Knockout

In yeast, mismatch repair is a multistep process that identifies and fixes incorrectly matched Watson-Crick base pairs in the genome. A mismatched DNA heteroduplex is first recognized by MutS α and MutS β heterodimeric complexes, which is composed of Msh2p/Msh3p and Msh2p/Msh6p, respectively. MutS heterodimers then interact with MutL homolog proteins Mlh1p and Pms1p to continue the recognition cascade and to recruit the correction machinery. (33, 34) Since the incorporation of mutagenic oligonucleotide DNA into the target genomic site produces a heteroduplex that is recognized by the MMR machinery, avoidance of this system is important to facilitate efficient oligo incorporation (35, 36).

Previous work has shown that *mlh1* and *msh2* knockouts can increase oligonucleotide incorporation in yeast. However, these knockouts result in varying recombination frequencies depending on the characteristics of the locus and oligonucleotide transformed. (33) **Figure 2.2 A, B, and C** show the efficiency of oligonucleotide allelic replacement using oligonucleotides with a single point mutation in the center of the target sequence when using strains with *mlh1* and *msh2* knockouts. Across all strains and loci, *mlh1* deletions increased median oligonucleotide incorporation from 3.9 to 11.0 fold, while *msh2* deletions increased median oligonucleotide incorporation from 2.9 to 9.2 fold. The double knockouts of *mlh1* and *msh2* in the CEN.PK113-7D and VTT C-68059a strains showed no significant effect in oligonucleotide recombination. We proceeded to use the strains with a single knockout in *mlh1* for further optimization, as this modification generally gave the highest increase of oligonucleotide incorporation across the different strains in 5 out of 7 tested loci.

Section Two

2.3 DNA Recombinase Overexpression

Single-stranded DNA recombinases are crucial for oligonucleotide recombineering in prokaryotic organisms, able to increase the rate of oligo incorporation by ~10,000-fold in some cases. (21, 29) Endogenous proteins crucial to the homologous recombination pathway in yeast also stimulate oligonucleotide incorporation. Overexpression of Rad51 and Rad54 proteins, as well as particular point mutants of Rad51, has previously been shown to affect oligonucleotide integration in yeast. (17, 37, 38)

Starting with the Δ mlh1 knockout strains, we constitutively expressed heterologous and endogenous DNA recombinases to examine their effect on oligonucleotide incorporation. In strain VL6-48, we expressed a yeast codon optimized version of the λ Red beta-protein, both with and without a nuclear localization signal tagged to the C-terminus of the protein. We were unable to observe a positive effect on oligonucleotide incorporation. (**Figure 1.2A**) This is contrary to work by Brachman et al. who observed a 9-fold increase in oligonucleotide incorporation at the CYC1 locus upon plasmid based overexpression of the λ Red beta-protein in a mismatch repair capable yeast strain. However, it is difficult to directly compare these results due to the different strain backgrounds and experimental methods utilized (e.g. different loci, oligonucleotide lengths, and expression systems). (16)

In a Δ mlh1 mutant background strain we overexpressed these proteins starting with RAD51, a DNA strand exchange protein crucial in homologous recombination. Lu et al. showed that a K342E mutant of RAD51 enhanced oligonucleotide recombination when compared to wild-type RAD51. This particular mutation was shown to affect the DNA-binding profile of RAD51 by removing its intrinsic inhibition to form filaments around single-stranded DNA in the presence of double-stranded DNA. (17, 37, 38) To examine if this effect held true in a mismatch repair deficient strain, we constitutively overexpressed both proteins genomically in the VL6-48 Δ mlh1 strain and found nearly a 3-fold increase in oligonucleotide incorporation in the RAD51(K342E) mutant as compared to the wild type RAD51. (**Figure 2.2 A**) Furthermore, RAD51(K342E) overexpression resulted in increased oligonucleotide incorporation compared to wild type

by up to 80.1 fold in VL6-48, 13.4 to 27 fold in CEN.PK113-7D, and 5.9 to 13.5 fold in VTT C-68059a. We additionally overexpressed RAD54, a helicase and chromatin remodeler, in strain CEN.PK113-7D Δ mlh1 in an attempt to further improve oligonucleotide incorporation efficiency. Overexpression of RAD54 in the Δ mlh1 background showed a 12 to 35 fold increase in oligonucleotide incorporation compared to wild-type. (**Figure 2.2 B**)

Overexpression of both RAD51(K342E) and RAD54 in an Δ mlh1 background across all three strains resulted in the highest increase in oligonucleotide recombination of 19.5 to 494.9 fold across all strains and loci (**Figure 2.2**) compared to wild-type. The variability of these fold increases are likely due to: i) the different promoters used for gene overexpression (i.e., TEF1 promoters in VL6-48 versus PGK1 and TPI1 promoters in CEN.PK113-7D and VTT C-68059a), ii) the site of genomic incorporation of the RAD54 overexpression cassette (i.e., the mlh1 locus for VL6-48 versus the pdc6 locus for CEN.PK113-7D and VTT C-68059a), and iii) the inherent variability of oligonucleotide incorporation across the loci tested. In the CEN.PK113-7D Δ mlh1 strain, we overexpressed both RAD51(K342E) and RAD54 from the same HO locus and from separate loci (i.e., HO and PDC6) to see if overexpression of both proteins in close proximity would affect oligonucleotide incorporation. Compared to wild-type, expression of the genes from separate loci resulted in a 77.7 to 179.3 fold increase in oligonucleotide incorporation, which was higher than the 28.3 to 50.6 fold increase in oligonucleotide incorporation when they were expressed at the same locus. (**Figure 2.2B**)

In the CEN.PK113-7D Δ mlh1 ho::RAD51(K342E) pdc6::RAD54 and VTT C-68059a Δ mlh1 ho::RAD51(K342E) pdc6::RAD54 strain backgrounds, msh2 was also knocked out, which provided no significant increase in oligonucleotide incorporation in the strains CEN.PK113-7D and VTT C-68059a. (**Figure 2.2 B and C**) We designate the most recombinogenic strains with the suffix, “.RR” to refer to their overexpression of RAD51 and RAD54. Hence, VL6-48 ho::RAD51(K342E) mlh1::RAD54, CEN.PK113-7D Δ mlh1 Δ ho::RAD51(K342E) Δ pdc6::RAD54, and VTT C-68059a Δ mlh1 Δ ho::RAD51(K342E) Δ pdc6::RAD54, are referred to as VL6-48.RR, CEN.PK.RR and VTT.RR, respectively.

2.4 External Factors Affecting Oligonucleotide Recombination

We further examined other factors that affect oligonucleotide recombination in yeast, including oligonucleotide length, amount of oligonucleotide transformed, and amount of homology to the site of integration. The intrinsic differences in transformation frequency between strains were also measured with the transformation of a plasmid having a centromeric origin of replication. Several studies have reported a variety of base values for oligonucleotide incorporation, depending on the locus. (15, 16, 34, 39) To obtain a range of oligonucleotide incorporation frequencies in each strain examined, we measured the frequency at three loci on separate chromosomes in each strain; ADE2, LEU2 and CAN1 for strain VL6-48, and URA3, LYP1, and CAN1 for strains CEN.PK113-7D and VTT C-68059a.

Optimal oligonucleotide length for oligonucleotide-mediated recombination in yeast has varied across different studies. Sherman et al. initially reported an optimum

oligonucleotide length of 50 base pairs, while several other reports have used 70 base pair oligonucleotides for generating changes.(15, 16, 39) In our most efficient strain, VL6-48.RR, we found that 90 base pair oligonucleotides gave the greatest increase in incorporation frequency when tested at the ADE2 locus. **(Figure 2.3A)** Other oligo-mediated recombination studies in *E. coli* have also found the 90mer to be optimal.(21) Hence 90mer oligonucleotides were used in all further recombination experiments. For VL6-48.RR, incorporation of 90mer oligonucleotides containing the noncoding sequence of the ADE2 gene was ~3.5-fold higher in frequency than oligos with the coding sequence. Annealed doubled stranded 90mer oligonucleotides gave the lowest incorporation frequency in this strain, resulting in recombination frequencies that were ~10-fold worse than the coding sequence oligonucleotides. **(Figure 2.3 B)** Furthermore, we titrated the amount of 90mer noncoding oligonucleotide used for each transformation and found that an oligonucleotide concentration of 2.5 μ M in the 400 μ l of cell electroporation mixture resulted in the highest transformation frequency. **(Figure 2. 3 C)**

The amount of homology to the locus of integration also affects the frequency of recombination. A premature stop codon was encoded in the 90mer oligonucleotides along with various other modifications of varying lengths such as mismatches, deletions and insertions. In the strain VL6-48.RR, mismatches were more tolerated than either insertions or deletions. **(Figure 2.3 D)** This is likely due to the preservation of the oligonucleotide length in the mismatched DNA. We observed that the oligonucleotide incorporation frequency decreases as the number of mismatches increases, consistent with previous observations in *E. coli*.(21) **(Figure 2.3 D)**

Additionally, we assayed transformation and survival frequencies via the transformation of a replicating plasmid to determine if oligonucleotide incorporation frequencies may be biased by intrinsic strain variability in DNA transformation. Indeed, we found that the wild-type VL6-48 strain had the highest survival and transformation frequency, which was consistent with its higher oligo-mediated recombination frequency. **(Figures 2.3 E and F)**

2.5 Locus Variability

We examined the variability of oligonucleotide recombination frequency across the genome in each strain studied. In strains CEN.PK.RR and VTT.RR three negatively selectable alleles were generated (using oligonucleotides containing premature stop codons), and in strain VL6-48.RR one negatively selectable allele and two positively selectable alleles were generated (using oligonucleotides correcting genomic premature stop codons). To designate which oligonucleotide was transformed, we denoted the sequences containing the coding or transcribed strand of the gene as COD, and the sequences containing the noncoding or non-transcribed strand of the gene as NCOD. **(Figure 2.4 A)** Taking the ratio of the most recombinogenic oligonucleotide and loci to the worst, we found a ~40 fold difference in recombination frequency across all alleles and strains tested, indicating that certain loci are more recombinogenic than others. **(Figure 2.4)** The highest recombination frequency, ~2%, was observed in strain VL6-48.RR at the ADE2 locus using a NCOD oligonucleotide **(Figure 2.4 B)**. CEN.PK.RR exhibited the next highest recombination frequencies at the loci tested **(Figure 2.4 C)** and the lowest recombination frequencies were observed in the industrial strain VTT.RR. **(Figure 2.4 D)**

Furthermore, we found that across all loci and strains there seems to be a lack of transcriptional bias towards rates of oligonucleotide-mediated recombination, potentially supporting a mechanism that oligonucleotides become incorporated during DNA replication. Unfortunately, because of an absence of empirical data regarding replication fork directionality in the loci examined, it is difficult to definitively conclude that either the leading or lagging strand is favored for oligonucleotide incorporation, however our data does not contradict a replication based model. We found that variability in strand bias in the examined loci was consistent between strains and could not be solely explained by transcriptional effects (**Figure 2.4**), which lend further mechanistic support for lagging strand incorporation of oligos at the DNA replication fork.(33, 39, 40)

2.6 Multiplex Modifications and Cycling

For the practical oligo-mediated directed evolution of yeast genomes, the ability to target multiple sites simultaneously is important. Furthermore, given that most of the single recombination frequencies observed in this study were below 1%, additional cycles of transformation may be required to enrich the population with modifications. To demonstrate multiplexing and cycling, we used strain CEN.PK.RR to target two negatively selectable genes CAN1 and LYP1 with the most efficient oligonucleotides to introduce premature stop codons (a NCOD oligonucleotide for CAN1 and a COD oligonucleotide for LYP1, **Figure 2.4 C**). We developed a cycling protocol allowing iterative transformation and recovery of the population with a turnaround of 12 to 16 hours. (**Figure 2.5 A**) We completed three cycles of oligonucleotide transformation and assayed the recombination frequency after each cycle. The transformed oligo pool

contained a mixture of 1.25 μM of each oligonucleotide yielding a total of 2.5 μM of oligonucleotides. We observed, as expected, that the frequency of incorporation increased as the number of cycles increased. **(Figure 2.5 B)** In contrast, the number of nonsense CAN1 or LYP1 mutants remained constant when no oligonucleotides were added to the transformation mix. **(Figure 2.5 D)** In cycle one, the individual oligonucleotide recombination frequencies were at most half of the individual recombination frequencies previously observed, which is likely due to the decreased concentration of the individual oligonucleotides used in the multiplexing experiment. **(Figure 2.5 B)** We observed that the co-incorporation frequency per cycle is less than the individual frequencies. **(Figure 2.5 C)**

2.7 Discussion and Future Considerations

Generating targeted diversity or allelic replacement in yeast is crucial for rational strain engineering. Here we have identified the combinatorial strain modifications and optimized the key parameters important for oligonucleotide recombination. Furthermore, YOGRE has achieved a single optimum oligonucleotide genomic incorporation frequency of 0.2 to 2.0% across all strains and loci tested without selection for the modification or stimulation for recombination by cleavage, which is at least 20 to 1000 fold higher than previously reported (14, 15, 17, 38, 39). In many cases, these levels of recombination are suitable for moderate screening efforts to isolate correct recombinants. Furthermore, this method is multiplexable and the population can be enriched for the desired mutations with additional cycles. Given the most and least recombinogenic strains and loci recombination frequencies, (VL6-48.RR with 2% at the ADE2 locus and VTT.RR with

0.15% at the URA3 locus), along with their respective survival rate post-electroporation (VL6-48 with ~1.0% and VTT C-68059a with 0.1%) and a constant number of 10^8 cells used per electroporation, we can estimate a range of potential library size of 10^2 to 10^5 genomic recombinants per locus per cycle. As several yeast library generation protocols use 10-fold more cells for electroporation, it is likely that these values could increase if more diversity is required.

The generation of genetic diversity in a rational, site-specific way would allow for higher fitness mutant libraries that could be screened for desirable phenotypes. For industrial strain evolution, using weak selections to relevant phenotypes, such as crude feedstock tolerance or cellular waste product resistance, the relevant modifications in the populations could be further enriched. Furthermore, easy generation of multiple alleles would allow for investigation of epistatic interactions between DNA regulatory regions and proteins. For example, pathways for desirable compounds could be optimized via promoter tuning using oligonucleotides.(20) Additionally, changes could be made to protein coding regions to generate mutant protein variants. (41)

Enrichment of modified cells, without a phenotypic selection, would greatly ease recovery of generated mutants using YOGI. Employment of site directed DNA double-strand breaks at unmodified genomic loci could be used to enrich for correctly modified cells. Unrepaired double strand breaks cause cell arrest and reduce cellular fitness, meaning that cells containing the cut sites will be at a disadvantage, allowing for a fitness advantage to cells without these cut sites. By using designable site-specific endonucleases, such as the RNA-guided endonuclease Cas9 or transcription activator like

endonucleases (TALENs), targeted to cleave wild-type unmodified alleles, a negative selection can be applied against these alleles to enrich the population for desired genotypes. (27)

The variability associated with YOGÉ across strains may stem from many factors. Individual differences between oligonucleotide recombination frequencies at different loci could be affected by protein occupancy at the locus, secondary structure of the donor oligonucleotide DNA, as well as replication dynamics proximal to the integration site. Furthermore, differences in strain oligonucleotide recombination profiles may relate to endogenous recombination proteome differences, along with reported transformation variability between strains. In addition we observed that the location of the integration of RAD51(K342E) and RAD54 expression cassettes can affect oligonucleotide recombination profiles. In the CEN.PK113-7D strain, integration of RAD51(K342E) and RAD54 expression cassettes in the same and at different safe harbor loci, result in different oligonucleotide recombination profiles, with the separate safe harbor loci integration method, resulting in a higher oligonucleotide recombination profile.

While the variability of oligonucleotide recombination between loci and strains may range by a factor of 10, cycling YOGÉ allows for step-by-step enrichment of desired genotypes allowing modifications to attain higher frequency with each cycle. In summary, YOGÉ is a useful method for integrating oligonucleotides with high frequency in yeast, with the ability to rapidly cycle the procedure. In combination with weak selections for desirable phenotypes or targeted endonuclease based DNA negative

selections, the targeted modified sequences or desired phenotypes could be enriched to easily screenable levels.

2.8 Figures

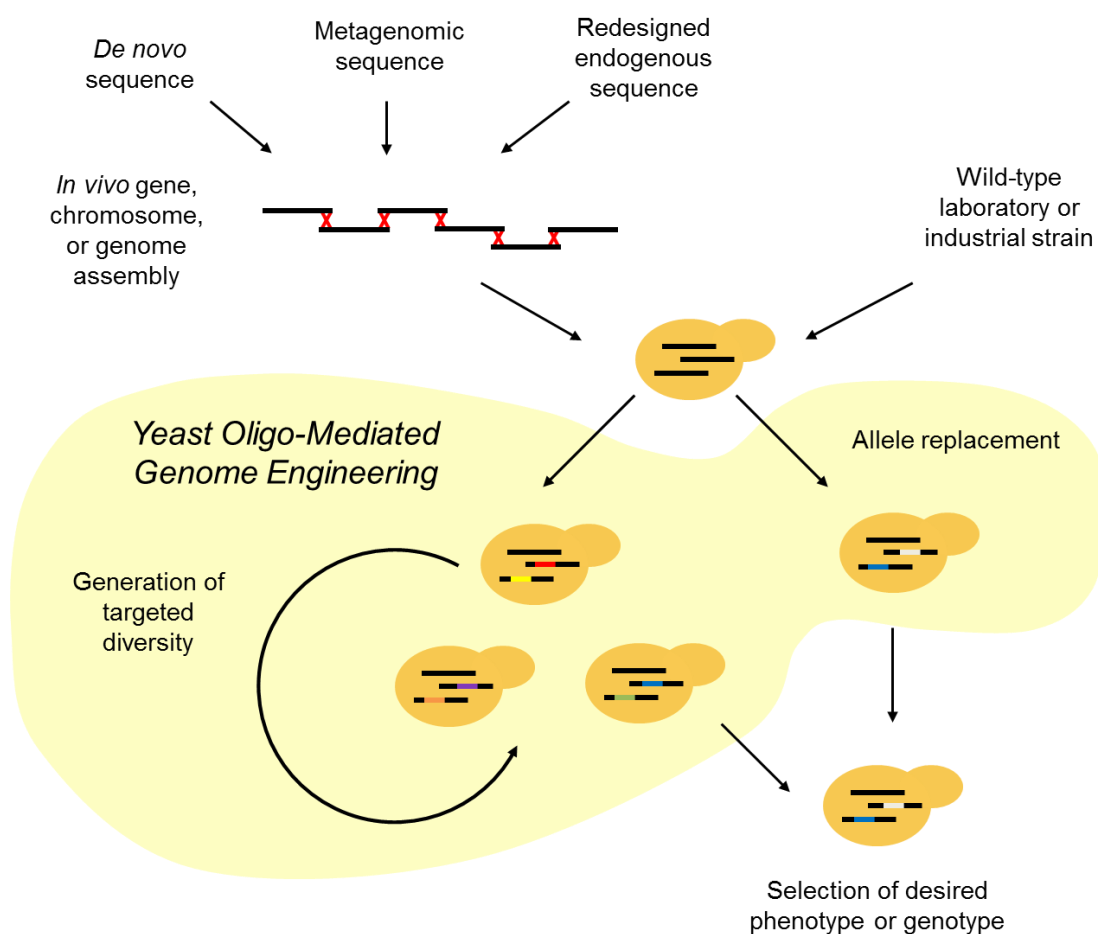


Figure 2.1. Diagram of Yeast Oligo-mediated Genome Engineering (YOG) applications. Synthetically assembled foreign DNA or natural genomes can be modified with YOG. Iterative rounds can be used to increase the frequency and/or diversity of the directed modifications. Finally, strains can be screened or selected to isolate desired genotypic or phenotypic traits.

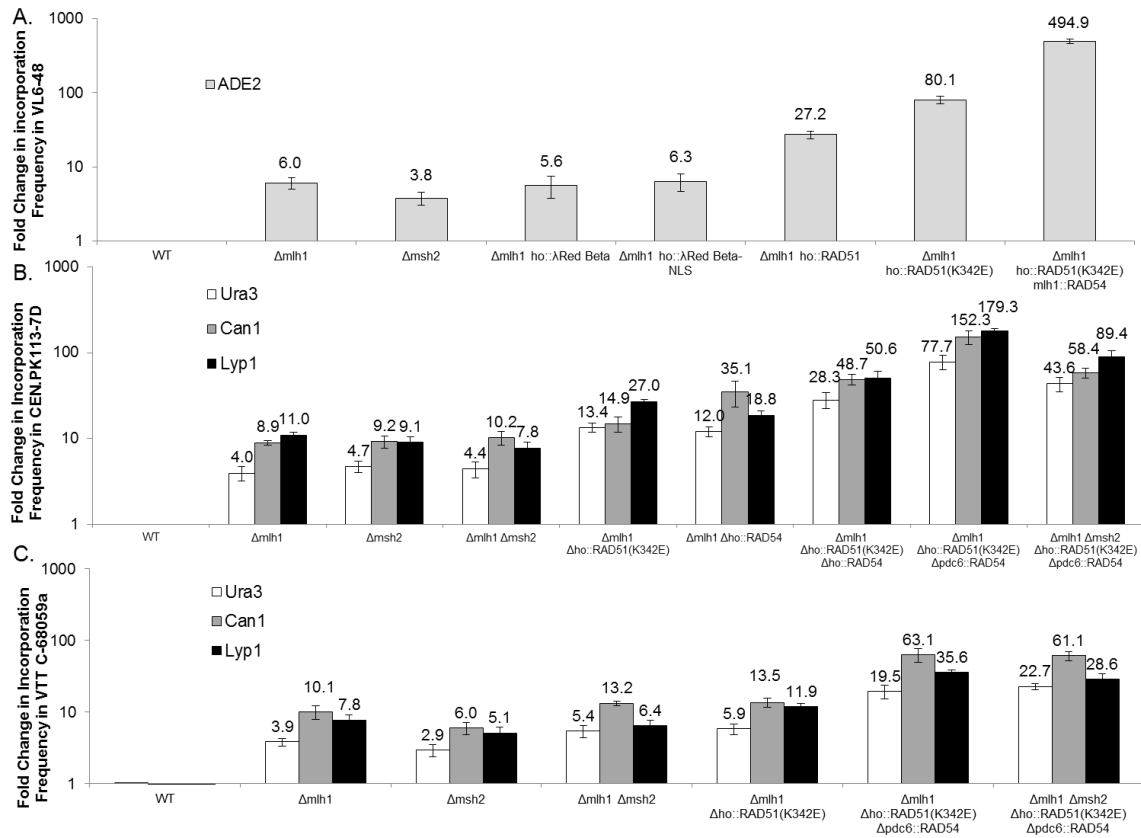


Figure 2.2 . Oligonucleotide recombination frequency across all strains and loci tested. Recombinations were completed using 90mer oligonucleotides with the optimal sense strand targeted. All recombinations were completed in triplicate, with values plotted to represent the mean values and error bars corresponding to the standard deviation. A. Oligonucleotide incorporation frequency in strain VL6-48 B. Oligonucleotide incorporation frequency in strain CEN.PK113-7D C. Oligonucleotide incorporation frequency in strain VTT C-68059a.

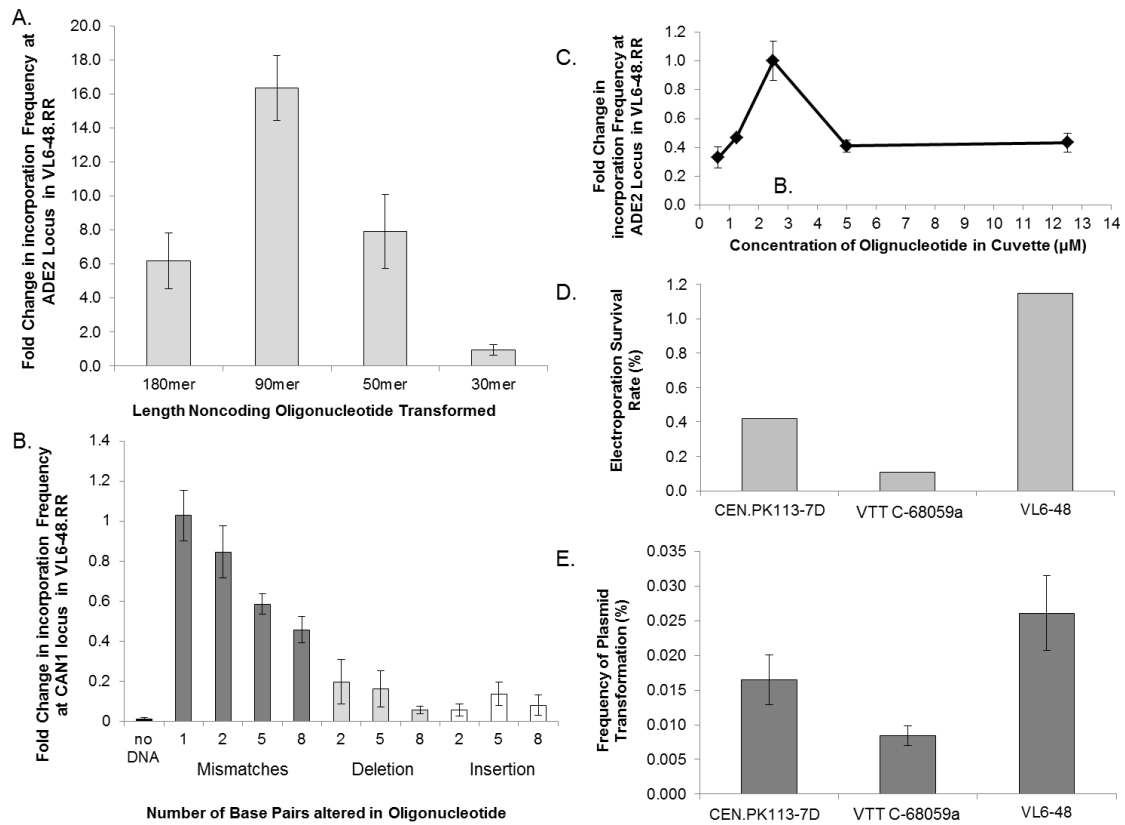


Figure 2.3 External factors affecting oligonucleotide recombination. All values represent mean values of three replicates with error bars representing the standard deviation. A. Effect of oligonucleotide length on recombination at the ADE2 locus in strain VL6-48.RR B. Effect of oligonucleotide concentration on recombination at the ADE2 locus in strain VL6-48.RR C. Effect of oligonucleotide strandedness on recombination at the ADE2 locus in strain VL6-48.RR D. Effect of mismatches, deletions and insertions on the oligonucleotide recombination frequency at the CAN1 locus in strain VL6-48.RR (the 1 mismatch, all deletion, and all insertion experiments were completed with 5 replicates) E. The survival rate of cells following electroporation across the three strains tested. F. Frequency of plasmid transformation in all three strains using a centromeric plasmid containing hygromycin resistance

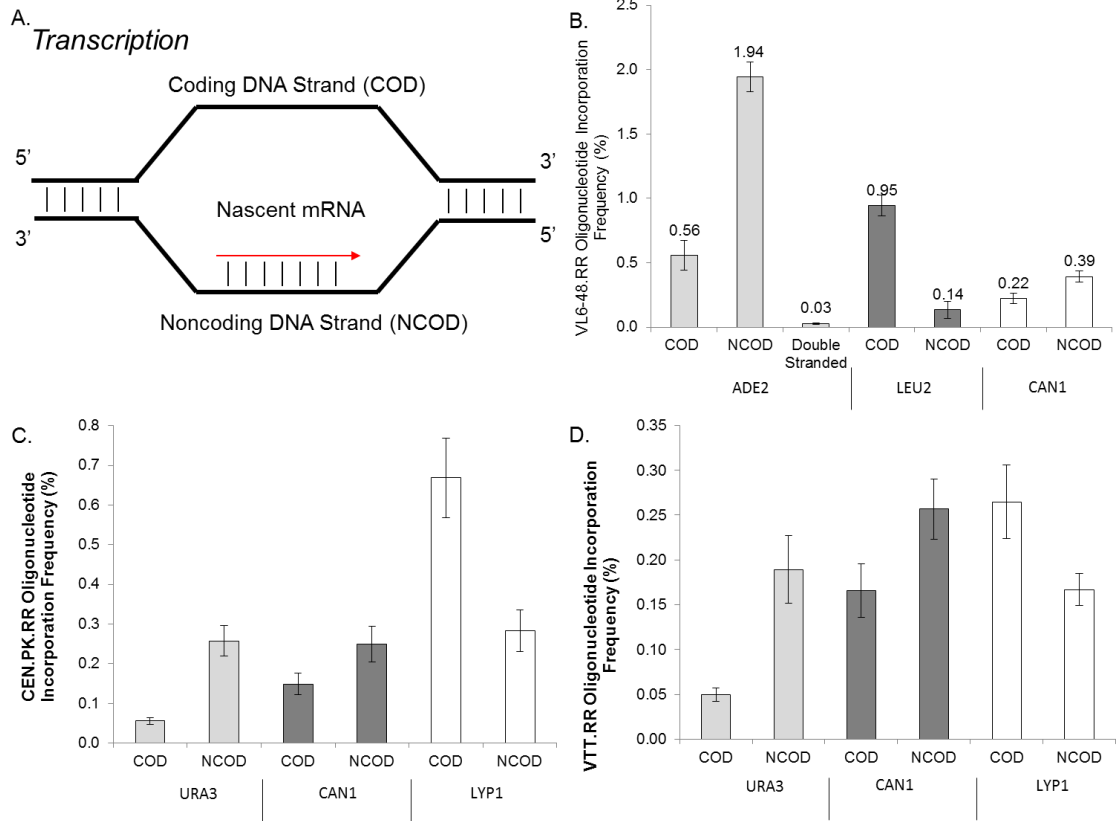


Figure 2.4 Variability in oligonucleotide recombination across DNA strands, loci and strains. All values represent the mean of three replicates with error bars representing the standard deviation. A. Diagram of the transcription fork with the coding and noncoding strands indicated. Oligonucleotides are named based on the sequence they correspond to in the transcription fork. B. Strand bias variability in oligonucleotide recombination across the ADE2, LEU2, and CAN1 loci of the VL6-48.RR strain. C. Strand bias variability in oligonucleotide recombination across the URA3, CAN1, and LYP1 loci of the CEN.PK.RR strain D. Strand bias variability in oligonucleotide recombination across the URA3, CAN1, and LYP1 loci of the VTT.RR strain.

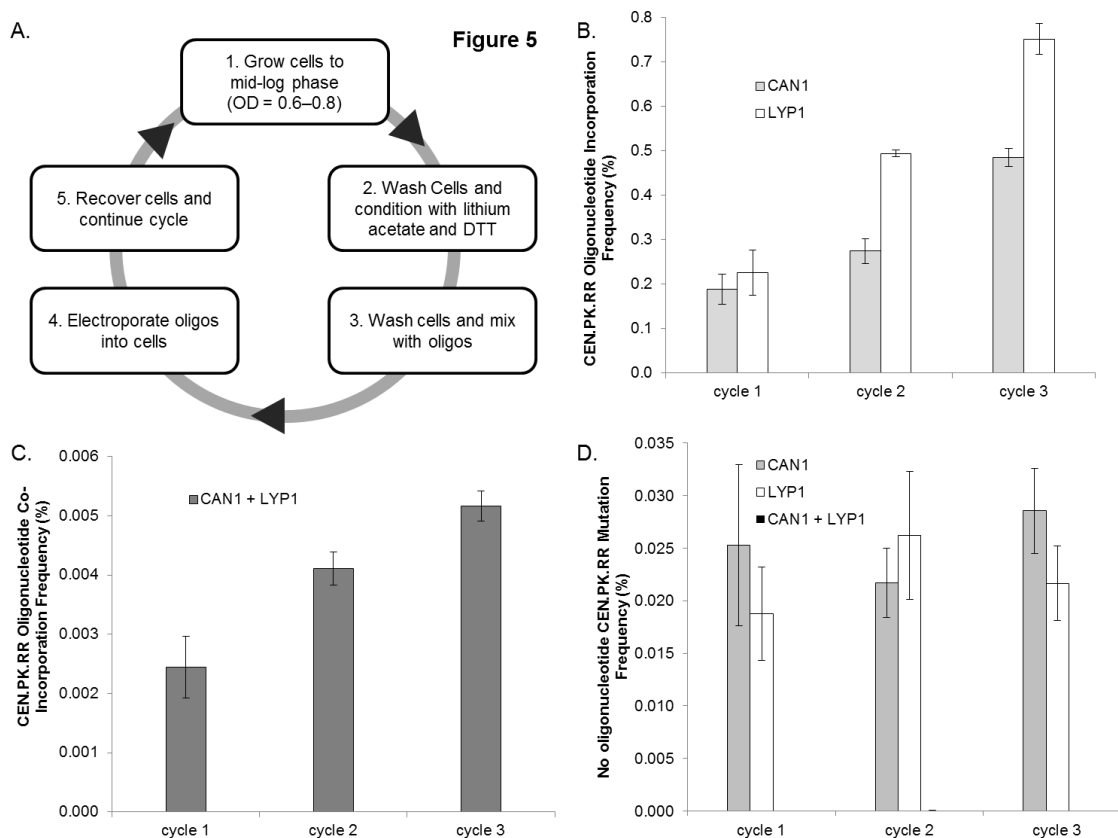


Figure 2.5 Cycling and multiplex modifications using YOG E in strain CEN.PK.RR at the CAN1 and LYP1 loci. All values represent the mean of three replicates with error bars representing the standard deviation. A. Diagram of YOG E electroporation cycling and recovery. B. Singleplex recombination frequencies for the CAN1 and LYP1 loci. C. Multiplex recombination frequency for both loci simultaneously. D. Rate of gene conversion without oligonucleotide added during electroporation, demonstrating that enrichment in loci modification is caused from transformation and not enrichment due to fitness advantage of any genotype.

CHAPTER 3 – Genome engineering in yeast using CRISPR-Cas9

3.1 Introduction

CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) and CRISPR-associated systems (Cas) immune systems in bacteria are of interest to the biotechnology community due to RNA-guided endonuclease activity.(25, 42) The Cas9 gene, from the type II bacterial CRISPR system of *Streptococcus pyogenes*, complexes with a designer genome targeting CRISPR guide RNA (gRNA) to determine the site-specificity of the DNA cutting activity. (25, 43) (**Figure 3.1 A**) It has been previously shown that Cas9 can function as an RNA-guided endonuclease in human cells.(44, 45) To express RNA without modifications added by the RNA polymerase II transcription system, RNA polymerase III regulatory elements have been used for transcription of functional gRNA in human cells. (44, 45)

To examine RNA-guided Cas9 nuclease activity we chose to design gRNAs to target the endogenous genomic negative selectable marker CAN1, a plasma membrane arginine permease, in haploid yeast cells and monitor mutation frequency at the locus. Nonsense mutations in CAN1 can be selected with media containing canavanine (a toxic arginine analog), which is only imported into cells containing a functional CAN1 gene. (46) Directed double-strand break at this locus increases mutation frequency due to errors that occur in the repair pathway. (46, 47) The double-strand break can be resolved either by homologous recombination or through error-prone non-homologous end joining. (14, 48, 49) To control for a potential genome-wide mutator phenotype, the mutation frequency of the non-targeted endogenous LYP1 gene, a lysine permease, was

monitored by selecting for *lyp1* mutants using a toxic lysine analog, thialysine.(50) The LYP1 and CAN1 genes are on separate chromosomes, and local mutation frequency in either locus should be independent unless a global mutator phenotype is present.

We further examined the effects of genomic CRISPR activity on single and double-stranded oligonucleotide transformation. It has previously been shown that induction of double-strand breaks near the oligonucleotide targeting site can increase recombination efficiency by as high as 4,000 fold. (14) To observe homologous recombination using CRISPR and donor DNA, a positive reporter system was chosen to avoid ambiguity from the negative reporter system, where of the source of mutations could be from erroneous double-strand break repair or donor DNA.

3.2 gRNA Expression and Genomic Target Identification

The RNA polymerase III regulatory elements used in this study were based on constructs used to express bacterial tRNA genes in yeast.(51, 52) Specifically, the SNR52 snoRNA promoter and the yeast tRNA gene SUP4 3' flanking sequence (as a terminator) were used to express gRNA. (**Figure 3.1 B**) This combination of promoter and terminator has been shown to produce transcripts with strings of U residues less than six ribonucleotides long, which is an important consideration as the structural component of the gRNA contains a string of four U residues.(44, 53, 54) The protospacer adjacent motif (PAM) sequence, a genomically encoded NGG nucleotide sequence directly 3' of the 20 base pair genome target, is a crucial design constraint to the site-specificity. The design of the gRNA structural component used in this study was based on the sequence used by Mali et al. for gRNA expression in human cells.(44) (**Figure 3.1 A and B**) The

20 base pairs of genome sequence complementary in the gRNA mimics the processed CRISPR RNA (crRNA) found in the natural host, *S. pyogenes*. In this organism, a 39-42 base pair sequence, containing a 20 base pair spacer-derived guide sequence and 19-22 base pair repeat-derived sequence, is processed by a trans-activating crRNA (tracrRNA) molecule and RNase III enzyme to form functional gRNA. The functional gRNA designs in human cells(45)(42)(41)(40)(39)(36)(34)(33)(31)(28), build upon the work of Jinek et al. who demonstrated *in vitro* that Cas9 requires both a base paired activating tracrRNA and the targeting crRNA, or a single chimeric guide RNA with features of each molecule. (25, 44, 45, 55) It has been reported that mismatches in any of the last 12 nt of the 20 nt crRNA against a target dsDNA can ablate Cas9 activity, while mismatches in the first 8nt have little effect. Therefore 23bp sequences of the form N(8)S(12)NGG for which the 12 bp “seed” sequence S(12) followed by an NGG is found nowhere else in the genome have the greatest chance of being specifically targetable by Cas9 However, partial activity of a NAG PAM sequence has also been observed for the Cas9 system(56), so that targets that may be unique genomic S(12)NGG occurrences but have one or more S(12)NAG occurrences may be less specific. Using these constraints we tabulated 645,392 genomic targets of maximal specificity (unique S(12)NGG with no S(12)NAG occurrences), and 108,493 genomic targets of lesser specificity (unique S(12)NGG but one or more S(12)NAG).

3.3 CRISPR-Cas directed CAN1 mutagenesis

The CAN1 locus was targeted with gRNA expression constructs on high-copy 2 μ plasmids. The Cas9 gene was placed under the Gal-L promoter, an attenuated version of

the strongly galactose-inducible Gal1 promoter, in a centromeric plasmid with ~1 copy per cell (**Figure 3.1 B**) and only induced with galactose during experiments.(57) We chose to examine Cas9 under an inducible promoter, at first, in order to limit the potential toxicity of CRISPR-Cas activity.

Two sites in the CAN1 gene were targeted. gRNA CAN1.Z was designed to direct endonuclease activity 58 base pairs downstream of the ATG start codon of the CAN1 gene, while the gRNA CAN1.Y genomic target site was located 207 base pairs downstream of the start codon. (**Figure 3.2 A**) Upon expression of Cas9 in strains also expressing gRNA, cell viability decreased to 78% and 89% with CAN1.Y and CAN1.Z, respectively, while strains containing only a single CRISPR-Cas component had viability near 100%. (**Figure 3.2 B**) The degree of toxicity also correlates with higher mutation frequency in the CAN1 locus, as CAN1.Y was the most toxic gRNA and had the highest CAN1 mutation frequency. Following galactose induction of Cas9, the mutation frequency in the CAN1 gene was 0.07 % and 0.01% with CAN1.Y and CAN1.Z gRNAs, respectively. (**Figure 3.2 C**) Furthermore, the mutation rate in the LYP1 gene remained relatively constant across all strains, suggesting that CRISPR-Cas is site specific in yeast and does not induce random mutations genome-wide. (**Figure 3. 2 C**) To further validate that mutations were caused by Cas9 and gRNA activity, the CAN1 gene from eight colonies were Sanger sequenced from the gRNA CAN1.Y/Cas9 and gRNA CAN1.Z/Cas9 canavanine resistant populations. Indeed, 8/8 colonies from the gRNA CAN1.Y/Cas9 canavanine resistant population and 7/8 colonies from the gRNA CAN1.Z/Cas9 canavanine resistant population were found to have frameshift CAN1

mutations directly upstream of their respective PAM sequences. (**Figure 3.2 D**) These mutations are proximal to the putative cleavage site of the Cas9 system, three base pairs upstream of the PAM sequence (25, 44, 45)

3.4 CRISPR-Cas stimulated homologous recombination with donor DNA and transient gRNA PCR cassette

As a test system for donor DNA homologous recombination, an allele containing a nonsense mutation of the ADE2 gene, a phosphoribosylaminoimidazole carboxylase essential for adenine biosynthesis, was targeted for repair. The *ade2-101* allele is common to many yeast lab strains and contains a premature stop codon at base 109 due to a G to T transversion. In this positively selectable reporter, mutations causing correction of the mutation would be rare without donor DNA. We chose to target this mutation using a 90mer oligonucleotide bearing the correct sequence of the ADE2 gene centered around the nonsense mutation. In yeast cells constitutively expressing Cas9 under the TEF1 promoter on a centromeric plasmid, either an ssDNA or dsDNA oligonucleotide containing these sequences were electroporated with either a transient PCR product of the gRNA cassette (containing the SNR52 promoter, gRNA sequence, and SUP4 3' flanking region) or salmon sperm DNA (as a control).

Two sites in the ADE2 gene were chosen to target with gRNAs based on the proximity of the *ade2-101* mutation. (Figure 3 A) Cells electroporated with gRNA cassettes and oligonucleotides showed higher rates of homologous recombination than compared to cells with only oligonucleotides. gRNA ADE2.Y had the largest effect on homologous recombination, improving single and double-stranded recombination rates 5-

fold and 130-fold, respectively. (Figure 3 B) The activity of ADE2.Z was not as high as ADE2.Y, potentially for several reasons. One reason may be that the cut site of the ADE2.Z was within the donor oligonucleotide, so interactions between the gRNA and oligonucleotide may have decreased genomic cutting ability. Also, a stretch of five U residues within the gRNA genomic targeting region may have resulted in decreased full-length ADE2.Z transcription efficiency RNA polymerase III due to premature transcription termination.

3.5 CRISPR-Cas stimulated homologous recombination with donor DNA and gRNA expression plasmid

In cells containing a centromeric plasmid constitutively expressing Cas9 under the TEF1 promoter, donor DNA was co-transformed with a high-copy 2 μ plasmid with and without gRNA CAN1.Y expression elements. Donor DNA was designed to recombine within the gRNA CAN1.Y genomic target using homology arms to the site. A double stranded 90mer oligonucleotide donor centered around the PAM sequence contained two base pair changes to mutate the PAM sequence and incorporate a premature TAG stop codon. A 1.4 kb KanMX cassette (conferring G418 resistance) was amplified with 50 base pair homology arms to the CAN1.Y target site and was also designed to disrupt the PAM sequence. Upon integration, both DNA donors would result in canavanine resistance. Selection of cells containing both gRNA.CAN1.Y and Cas9 plasmids resulted in a reduction of transformation frequency as compared to cells with a Cas9 plasmid and an empty vector. **(Figure 3.3 C)** This is likely due to toxicity of Cas9 DNA cleavage.

Furthermore, co-transformation of donor DNA with the gRNA.CAN1Y expression plasmid increased the transformation frequency as compared to a no donor DNA control (**Figure 3.3 C**) Colonies containing both plasmids were then replica plated to canavanine media as well as rich media with G418 (to select for the KanMX integration event). Interestingly, the vast majority (near ~100%) of the cells that received donor DNA selected for the gRNA CAN.1 Y and Cas9 plasmids were canavanine resistant and the same proportion were G418 resistant in the KanMX donor DNA co-transformation. (**Figure 3.3 D**) A small amount of transformants (an average of 2 colonies per replicate) resulted when gRNA CAN1.Y plasmid was transformed without donor DNA, all of which were canavanine sensitive. In the inducible Cas9 system, which lacked mutagenic donor DNA, canavanine resistant mutants arose with a frequency of 0.07%. Hence, in the no donor DNA control of this experiment there were too few transformants to likely observe such a low frequency. In cells co-transformed with an oligonucleotide containing a premature stop codon, to ensure that donor oligonucleotide recombination was the cause of canavanine resistance, we sequenced 8 colonies of the canavanine resistant population. Indeed, all 8 of these cells contained PAM sequence mutation and premature stop codon. This data suggests that under strong constitutive expression of Cas9, a genomically targeted gRNA on a plasmid can both stimulate recombination of donor DNA and select against wild type sequences with high frequency.

3.6 Materials and Methods

3.6.1 Strains and Media.

The *Saccharomyces cerevisiae* strain used in the CAN1 mutagenesis analysis of the CRISPR system and the gRNA plasmid/donor DNA transformation in Cas9 expressing cells was BY4733 (MATa *his3Δ200 trp1Δ63 leu2Δ0 met15Δ0 ura3Δ0*), which was a kind gift from Fred Winston. Parental BY4733 was grown in YPAD before transformation and then propagated in the appropriate synthetic complete (SC) media minus the auxotrophic compound complemented by the plasmids. Strain VL6-48 (MATα, *his3Δ200, trp1Δ1, ura3-52, ade2-101, lys2, psi^o, cir^o*) was used for the homologous recombination experiments, due to its native *ade2-101* premature stop codon. VL6-48 was purchased from ATCC (MYA-3666). Plasmids p415-Gal-L and p426-Gal1 used in this study and were a kind gift from Fred Winston. (57)

3.6.2 Plasmid Construction.

The Cas9 gene was a codon optimized version originally constructed for expression in human cells.(44) This gene was C-terminally tagged with a SV40 nuclear localization signal. The p415 Gal-L and p414 TEF1 plasmids were each cut with XhoI and XmaI and the backbone containing the promoter and terminator was gel purified. Cas9 was PCR amplified from a TOPO-TA vector with 20 base pair extended 5' and 3' regions identical to the promoter and terminator of the destined plasmid backbone (either p415 Gal-Lp or p414 TEF1p). The PCR amplified Cas9 was Gibson assembled into the vector using the NEB Gibson Assembly kit. For the gRNA expression plasmids, the

p426-Gal1 plasmid was cut with XhoI and SacI to remove the Gal1 promoter, and the backbone was gel purified. The gRNA expression cassette containing the SNR52 promoter, the gRNA and SUP4 3' flanking sequence were assembled by two rounds of PCR using Phusion 2X HF Master Mix. The outer two primers contained 20 base pair extended 5' and 3' regions identical to the p426-Gal1 plasmid backbone at the cut sites. In the first round of PCR contained all primers at 10 nM. The second round of PCR was a 50 µl reaction containing a 2 µl of 10-fold dilution of the first round product with the outer primers at concentrations of 10 nM. For the CAN1 experiments, the gRNA PCR products were Gibson assembled into the cut p426 plasmid. The KanMX sequence was PCR amplified with 50 base pair homology arms the CAN1 locus from the pFA6a-KanMX6 plasmid, commonly used for gene knockout in yeast.

3.6.3 Transformation of plasmids for CAN1 mutagenesis.

Transformation of plasmids (200 ng per transformation) was carried out using a standard lithium acetate transformation method. (58) After transformation, cells were plated on selective media (SC-uracil and leucine, SC-tryptophan, or SC-uracil and tryptophan) and allowed grow for 2 days until colonies appeared. All oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA, USA). Double-stranded oligonucleotides were generated by annealing equimolar amounts of single-stranded oligonucleotides by first denaturing the mixture at 100°C for 5 minutes and then allowing cool to 25°C with a ramp of 0.1°C per second.

3.6.4 Electroporation of Transient gRNA cassette and oligonucleotides.

gRNA cassette and donor single-stranded or double-stranded oligonucleotide were transformed via electroporation as follows. Cultures were grown to saturation overnight in SC-tryptophan. The next morning, two 10 mL cultures were inoculated in liquid SC-tryptophan to and OD600 = 0.3. Inoculated cells were grown in roller drum at 30°C until OD600=1.8 after 5 hours. Cells were collected via centrifugation at 2250 x g for 3 minutes and the media removed. The cell pellet was washed once by 10 mL ice-cold water and once by 10 mL of ice-cold electroporation buffer (1 M Sorbitol / 1 mM CaCl₂). The cells were conditioned by re-suspending the cell pellet in 2 mL 500 mM LiAc/10 mM DTT and placed in roller drum for 30 minutes at 30 °C. Conditioned cells were collected by centrifugation, and washed once by 10 mL ice-cold electroporation buffer. The cell pellet was re-suspended to a final volume of 6 mL in electroporation Buffer. 400 µl of cells were used per electroporation with 1 nmol of oligonucleotide and 1 µg gRNA cassette (or 1 µg of salmon sperm DNA as a control). Cells were electroporated at 2.5 kV, 25 µF, 200Ω. Electroporated cells were transferred from each cuvette into 7 mL of 1:1 mix of 1 M sorbitol /YPAD media. The cells were incubated in a roller drum at 30°C for 12 hours. Approximately 10⁶-10⁷ cells were plated on selective media and cells were diluted appropriately on rich media. The ratio of colony count on selective plates over rich plates was used as a measure of correction frequency. Experiments were completed in quadruplicate.

3.6.5 Transformation of gRNA plasmid with donor DNA into Cas9 expressing cells.

500 ng of either empty p426 or p426 containing a gRNA CAN1.Y expression cassette were transformed into BY4741 cells constitutively expressing Cas9 under the TEF1 promoter in a p414 plasmid backbone. The p426 plasmid was co-transformed with either 1 nmol of double stranded CAN1.Y oligonucleotide, 5 µg of KanMX cassette and 50 µg of salmon sperm DNA, or just 50 µg salmon sperm DNA, using a standard lithium acetate transformation method. (58) Cells were plated without dilution on SC without uracil and tryptophan and 10^{-5} dilutions onto YPAD and allowed to recover for 2 days before the selective plates were replica plated to canavanine plates and YPAD plates with 100 µg/ml G418 antibiotic (Geneticin G418, purchased from Teknova) Transformation frequency was calculated by the ratio of number cells that recover on SC without uracil and tryptophan divided by the number on rich nonselective media. No donor DNA experiments were completed with three replicates and the transformations containing donor DNA were completed with six replicates.

3.6.6 Galactose induction of Cas9.

Cells were grown in 5mL SC dropout without leucine and uracil media containing 2% glucose to saturation, washed twice in water, and then inoculated to an OD = 0.3 in SC dropout without leucine and uracil media containing 2% galactose and 1% raffinose. Cells were allowed to grow 16 hours before plating on YPAD, SC-arginine plates containing 60 µg/mL L-canavanine (Sigma), and SC-lysine containing 100 µg/mL thialysine (S-2-aminoethyl-L-cysteine, Sigma). 10^7 to 10^8 cells were plated on canavanine and thialysine containing media and cells were diluted appropriately on rich media. The

ratio of colony count on canavanine or thialysine plates divided by the colony count on rich media plates for each culture was used as a measure of mutation frequency. Experiments were completed in quadruplicate. Sequence alignments of Sanger sequence files were completed using Lasergene Seqman Pro Software.(59)

3.6.7 Toxicity of CRISPR system.

To analyze the toxicity associated with of the Cas9 protein and gRNA, strains were grown in 5mL SC dropout media containing 2% glucose to an OD = 2.0, washed twice in 5 mL water, and then equal amounts of cells plated in on both YPAD and YPA Gal (2% galactose and 1% raffinose) and allowed to grow at 30°C for two days. Experiments were completed in quadruplicate.

3.6.8 Identification of CRISPR-Cas targets in yeast genome.

Both strands of the complete the *S. cerevisiae* S288c genome (version R64-1-1, GenBank assembly GCA_000146045.2) were searched for sequences of the form N(21)GG that did not contain any string of 6 or more Ts, yielding 924177 candidate Cas9 target sequences. Sequences containing 6 Ts in a row were excluded as they can cause termination of RNA polymerase III transcripts.(51, 53) The yeast genome was then searched for other genomic occurrences of the S(12) “seed” sequences (see text) within these targets followed by NGG using bowtie version 1 release 0.12.8 using parameters -l 15 -v 0 -k 2. Specifically, for each S(12), the four sequences S(12)AGG, S(12)CGG, S(12)GGG, and S(12)TGG were considered input “reads” to be mapped against the genome, and any candidate for which more than one match was found among these four

reads was rejected. (60) All candidates passing this filter meet the Cas9 specificity conditions described in the text. A second round of seed checks was then performed using this same method looking for S(12)NAG matches. Here, targets that returned no matches for any of their four NAG reads were tagged as NoNAG targets to denote their greater specificity.

3.7 Discussion and Future Consideration

CRISPR-Cas has great potential as a foundational tool for genome engineering in *S. cerevisiae* due to the user-designated site-specificity of Cas9 endonuclease activity and the simplicity of gRNA construction. Yeast genome engineering methods utilizing site-specific endonucleases could also benefit greatly from CRISPR systems. The *delitto perfetto* method for genomic oligonucleotide recombination in yeast, as described by Storici et al., uses an induced double-strand break near the site of oligonucleotide recombination to obtain recombination frequencies of up to 20%. (14, 49) This technique requires initial targeted insertion of a CORE cassette containing a selectable marker with an I-SceI homing endonuclease site and a separate inducible I-SceI gene. Using a similar approach, CRISPR-Cas could greatly simplify this method by removing the initial step requiring I-SceI endonuclease site integration. Furthermore, our experiments using a transient PCR product of the gRNA cassette show that the site-specificity of the endonuclease could be directed easily using a non-integrating PCR product. More optimization, however, is required to attain these high frequencies of oligonucleotide recombination with a transient gRNA CRISPR system.

Transformation of donor DNA into cells selected for the presence of CRISPR

components on plasmids is a closer comparison to the *delitto perfetto* system, as site-specific DNA cleavage is more likely in all cells than on a transient PCR cassette. In this experiment, we saw recombination frequency near 100%, without selection for the integrated DNA. Removal of the PAM sequence in the donor DNA allowed recombinant cells to be protected from CRISPR-Cas activity and likely the associated toxicity. While we did not test this hypothesis, mutations in the 12 base pair “seed” sequence, crucial for site specificity, may result in similar protection. Given the ease of designing site-specificity for CRISPR-Cas targets with the high recombination frequency achieved for both a 90mer double stranded oligonucleotide and 1.4 kb double stranded DNA cassette, we believe this method could be extremely valuable for engineering yeast genomes without the need of selectable markers attached to the integrated DNA.

In addition, the mutagenic capabilities of CRISPR-Cas, as demonstrated by the CAN1 mutagenesis experiments, highlight the potential for the use of CRISPR-Cas to make targeted knockouts. Further improvement in percent mutagenized with an inducible system is needed before this would be a practical application, as the described induced Cas9 mutagenesis experiment only produced knockout rates of at most 0.07%. However, we have shown that if a mutagenic donor DNA to knockout the target gene is provided for homologous recombination, that this frequency can be dramatically boosted.

One exciting finding of this study was the use of a transient gRNA cassette to stimulate homologous recombination. Because no plasmids or selectable markers were needed to produce functional gRNA in Cas9 expressing cells, many gRNAs can thus be easily synthesized and combinatorially transformed. This could allow for genomic

targeting of many loci at once. Moreover, studies in several organisms have shown that multiple genomic targets are possible with the Cas9 system.(44, 61–63) Indeed in the future, we plan to target multiple sites with gRNAs simultaneously, either on a plasmid or in a transient PCR product, (potentially in an array form) and examine the possibility of CRISPR-Cas directed multiplex genome engineering towards engineering whole metabolic pathways and large gene-networks.

3.8 Figures

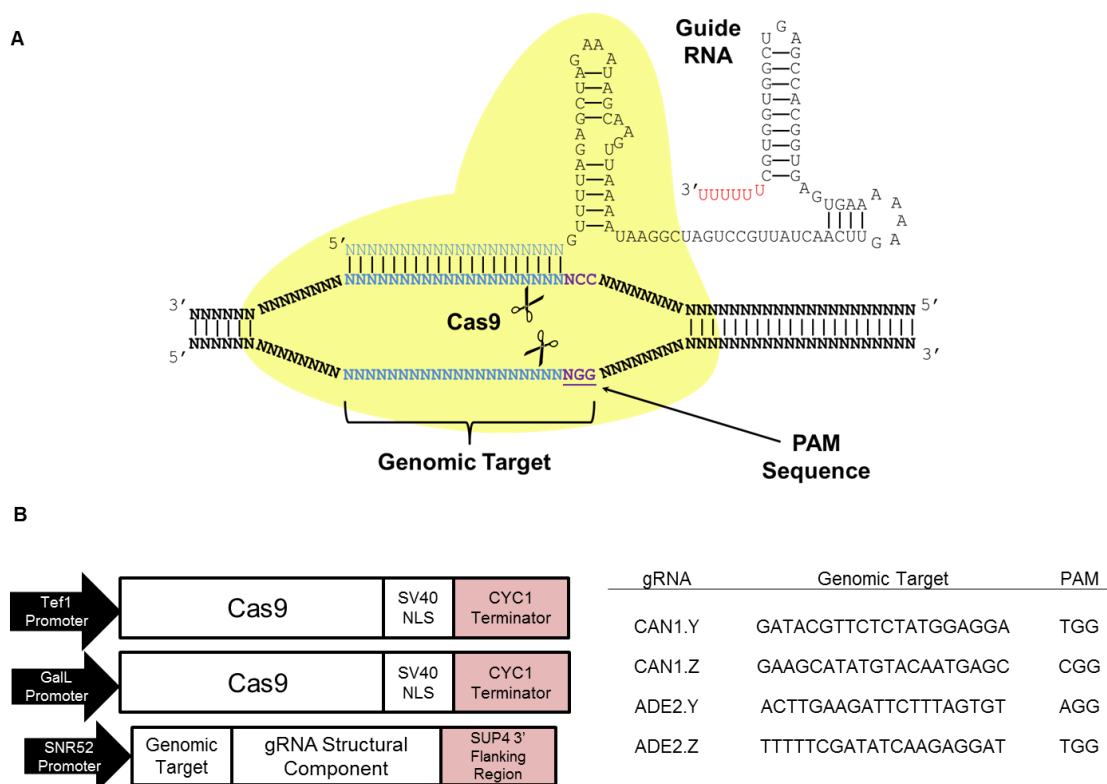


Figure 3.1 Diagram of Cas9 complex and schematic of genetic constructs A. Illustration of Cas9 protein interacting with Guiding RNA (gRNA) to direct endonuclease activity proximal to the protospacer-associated motif (PAM sequence). B. Design of the Cas9 and gRNA constructs. Cas9 gene contained a SV40 nuclear localization signal and was expressed under the Gal-L inducible promoter in CAN1 experiments and the Tef1 constitutive promoter in ADE2 experiments. The gRNA was expressed under the snoRNA SNR52 promoter and contained a terminator from the 3' region of the yeast SUP4 gene. CAN1.Y and CAN1.Z were targeted to different loci in the CAN1 gene, while ADE2.Y and ADE2.Z were targeted to different loci in the ADE2 gene

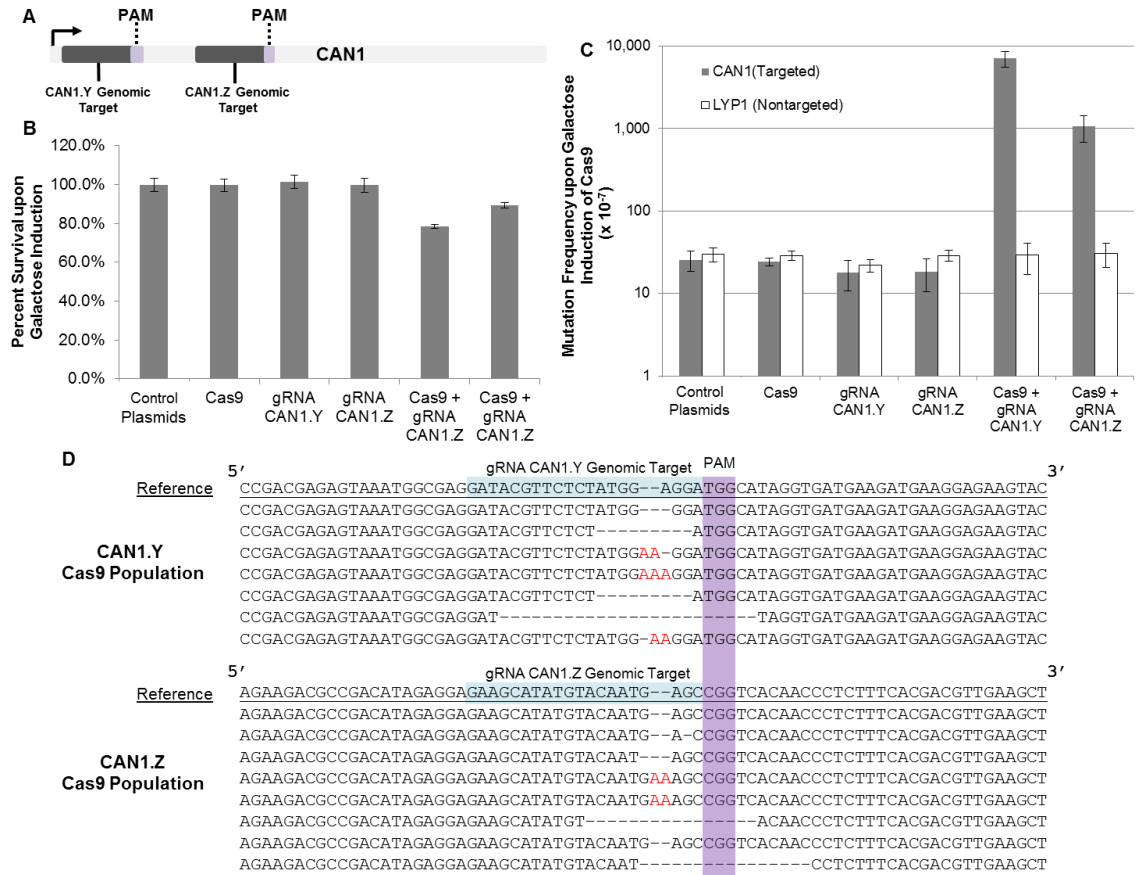


Figure 3.2 CRISPR mediated genomic mutagenesis. A. Schematic of the CAN1.Y and CAN1.Z genomic target sequences and respective PAM sequences in the CAN1 gene. **B.** Percent survival of yeast strains expressing different combinations of Cas9 and gRNA. Only combination of Cas9 and gRNA caused decrease in cell viability. Error bars represent standard deviation between four experiments. **C.** CAN1 and LYP1 mutation frequency of yeast strains expressing different combinations of Cas9 and gRNA upon galactose induction of Cas9. Mutation frequency of CAN1 (the gRNA targeted gene) is elevated in strains expressing Cas9 and gRNA, while LYP1 (not targeted by gRNA) remained constant. Error bars represent standard deviation between four experiments. **D.** Alignments of CAN1 gene from canavanine resistant colonies post-galactose induction of Cas9 in gRNA expressing strains. The underlined sequence is of the WT reference CAN1 gene, while the following eight sequences are from colonies from the population. The PAM sequence is highlighted in purple, while the gRNA guiding sequence is highlighted in blue. Nonsense mutations observed are largely deletions or insertions 5' to the PAM sequence.

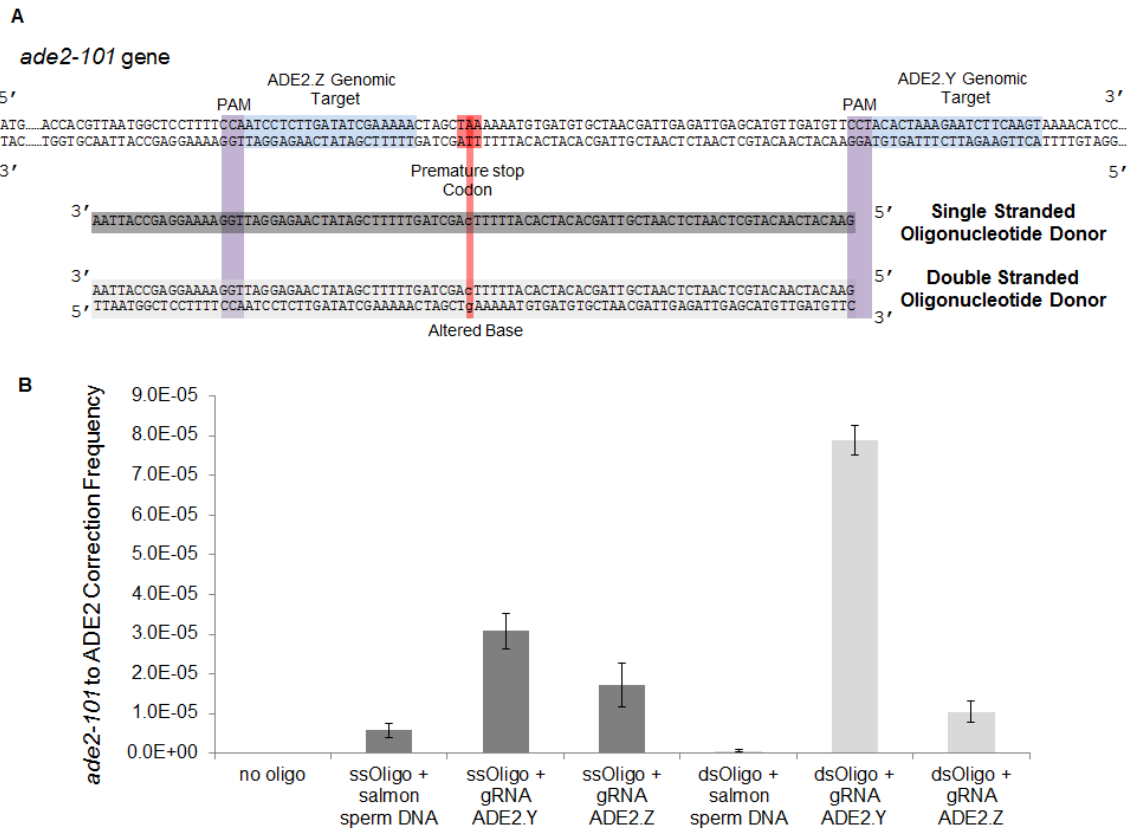


Figure 3.3 Transient Transformation of gRNA cassette and oligonucleotide recombination in Cas9 constitutively expressing strains. A. Diagram of ADE2 gene in the 5' to 3' direction displaying the gRNA genomic target regions and their respective PAM sequences. The sequence location of the single or double stranded oligonucleotide that will repair the premature stop codon is also displayed. The single stranded oligonucleotide is from the noncoding strand of the ADE2 gene, while the double stranded oligonucleotide is composed of annealed noncoding and coding strands. The oligonucleotide was designed to centered around the mutant base pair. The gRNA sequences were designed based on the closest PAM sequences to the premature stop codon. B. 90mer single and double stranded oligonucleotides were electroporated into yeast expressing Cas9 under the TEF1 constitutive promoter, with and without a gRNA cassette. The gRNA cassette is a PCR product consisting of the SNR52 promoter, the genomic target, the guiding RNA scaffold and the SUP4 3' flanking sequence. Electroporation of Oligonucleotide and ADE.Y gRNA cassette increases oligonucleotide incorporation up to 5 fold for single stranded oligonucleotide and 130 fold for double stranded oligonucleotide. Values are averages from four experiments and error bars represent standard deviation from the experiments

CHAPTER 4 – Optimization and Application of Cas9-mediated genome engineering

4.1 Introduction

Genome engineering in yeast using Cas9 has clear advantages. The biggest advantage is the high frequency oligonucleotide and gene cassette insertion. To explore if the initially described strain modifications used in the YOGES strain, will assist in oligonucleotide recombination with genomic Cas9 cleavage, I transformed oligonucleotides and a gRNA PCR product and to direct Cas9 cleavage to three loci in the described strain background. In a Cas9 expressing strain I compared oligonucleotide recombination between either solely single or double stranded DNA transformation, or single or double stranded DNA transformed with gRNA PCR products. Unfortunately there was no significant increase in oligonucleotide recombination when Cas9 was active at those loci. (**Figure 4.1A**) This is possibly due to the recombination pathway in the YOGES strain being depended on RAD51 and RAD54 strand invasion activity, while the recombination pathway for the Cas9 mediated method being dependent on a RAD52 dominated pathway. This is supported by models showing that RAD52 is most crucial in double stranded DNA recombination, and that RAD51 and RAD54 are important for strand invasion, as well as show activity at replication forks.(64, 65) From these results, as well as from constraints in library size from the YOGES method, I have further optimized of the multiplexability of the Cas9 genome engineering technique. One advantage to the Cas9 only method of genome engineering is that it doesn't require strain modification and would be potentially portable to many strains of yeast.

4.2 Multiplex Oligonucleotide Genome Editing via Cas9 with cycling

We have so far focused in optimization of oligonucleotides recombination using Cas9 in yeast. Using double stranded oligonucleotides, we aimed to optimize the method of transformation using a gRNA plasmid co-transformed with the donor DNA into strains constitutively expressing Cas9. To optimize the multiplex Cas9 mediated genome engineering protocol, I targeted two genes: the CAN1 gene, which was used in the preliminary Cas9 experiments and a broken KanMX gene, a geneticin resistance gene, containing a premature stop codon. I transformed double stranded oligonucleotides to inactivate the CAN1 gene, to allow growth on canavanine media, and activation of the KanMX gene to allow growth on geneticin containing media. Previous studies have found appropriate double stranded oligonucleotide concentrations for similar transformation conditions, so I sought to optimize the gRNA plasmid transformation amount. The more plasmid added, the more cotransformants and library sizes were possible. With 5 µg of each plasmid, a co-recombination rate of each oligonucleotide was approximately 5% with the entire population having each individual change at an individual frequency of 30% (**Figure 4.1B**) Using this method and parameters, we developed a cycling protocol so that high frequencies of combinatorial changes would be possible. To accomplish this we chose to use a gRNA plasmid containing the positive-negative selectable marker, URA3. This marker allows for initial selection for the plasmid, ensuring high frequency genomic changes to have occurred, then removal of selection to allow plasmid loss, followed by selection against the plasmid. This procedure would generate a specifically altered population that is ready to be transformed again

with gRNA plasmids and donor DNA double stranded oligonucleotides. **Figure 4.1C** shows a general schematic of the cycling procedure using Cas9 for genome engineering in yeast. Using this cycling procedure, we targeted four genes to impart mutations important in thermotolerance in yeast (as described in a subsequent section) two cycles were completed with and assayed the population for directed mutation using allele-specific PCR. **Figure 4.1D** shows as the cycle continue, the number of mutants increase in the population greatly, and even after one cycle nearly 80% of the population has at least one modification.

4.3 Optimization of CRISPR-Cas9 for oligonucleotide replacement and Donor DNA Delivery

By testing several concentrations of oligonucleotides and plasmids to co-transform into yeast, we were able to optimize the library size possible using Lithium Acetate transformation methods. (**Figure 4.2 A and B**) To do this we used the previously described CAN1 negative reported system and a donor DNA that imparts a premature stop codon. **Figure 4.2 A** shows the effect of increasing the amount of double strand oligonucleotide donor. Interestingly there was little effect with increasing amount of donor DNA. **Figure 4.2 B** shows the effect of increasing the amount of gRNA and Cas9 containing plasmid. 20 μ g of plasmid seems to show the highest transformation frequency corresponding to 1:1000 transformed cells receiving an appropriate modification. Typical transformations use 10^8 cells, meaning the potential library size for this method would be up to 10^5 cells.

4.4 Strain Engineering of Industrially Relevant Phenotypes

4.4.1 Thermotolerance

Thermotolerance is an industrially desirable phenotype in *S. cerevisiae*, as fermentation processes that can operate at higher temperatures are less likely to be contaminated and can reduce cooling costs in production plants. Several gene knockouts and alleles in yeast have been implicated in thermotolerant phenotypes. Specifically the genes SCH9, TFS1, and RAS2 are important regulators of stress responses and knockout of these genes can reduce overwhelming stress responses that may be detrimental to cell viability during long exposures to high temperatures. (66–68) Additionally, their absence may shift stress response pathways potentially so that cells may be better prepared to survive heat shock. Cells lacking these genes individually survive significantly longer under 55°C heat shock conditions. Additionally, affecting ubiquitination has been shown to effect thermotolerance in yeast. Specifically, a point mutation that affects a phosphorylation site of the ubiquitin-conjugating enzyme UBC1, disallowing phosphorylation has been shown to increase thermotolerance to a 2 hour 45°C heat shock.(67, 68) This mutation potentially recapitulates a dephosphorylation event that occurs under heat stress, so cells containing this mutation may be preconditioned to survive heat stress. (67)

Using Cas9 mediated genome editing cycling procedure described previously, we have generated nine mutants using the cycling procedure. **Figure 4.1 D** shows the proportion of modified cells per cycle with the number of mutations present in the population. We assayed the thermotolerant properties of these 9 generated strains and

show that there may be some epistatic effects between mutations in the subset of mutants generated. **Figure 4.3** shows the subset of mutants currently isolated and their ability to survive a 2-hour heat shock at 55°C. The generation of these alleles, and the potential epistasis that is resulting in phenotypic variation amongst these mutants, shows the potential of utilization of multiplexed Cas9-mediated genome engineering in yeast.

4.5 Figures

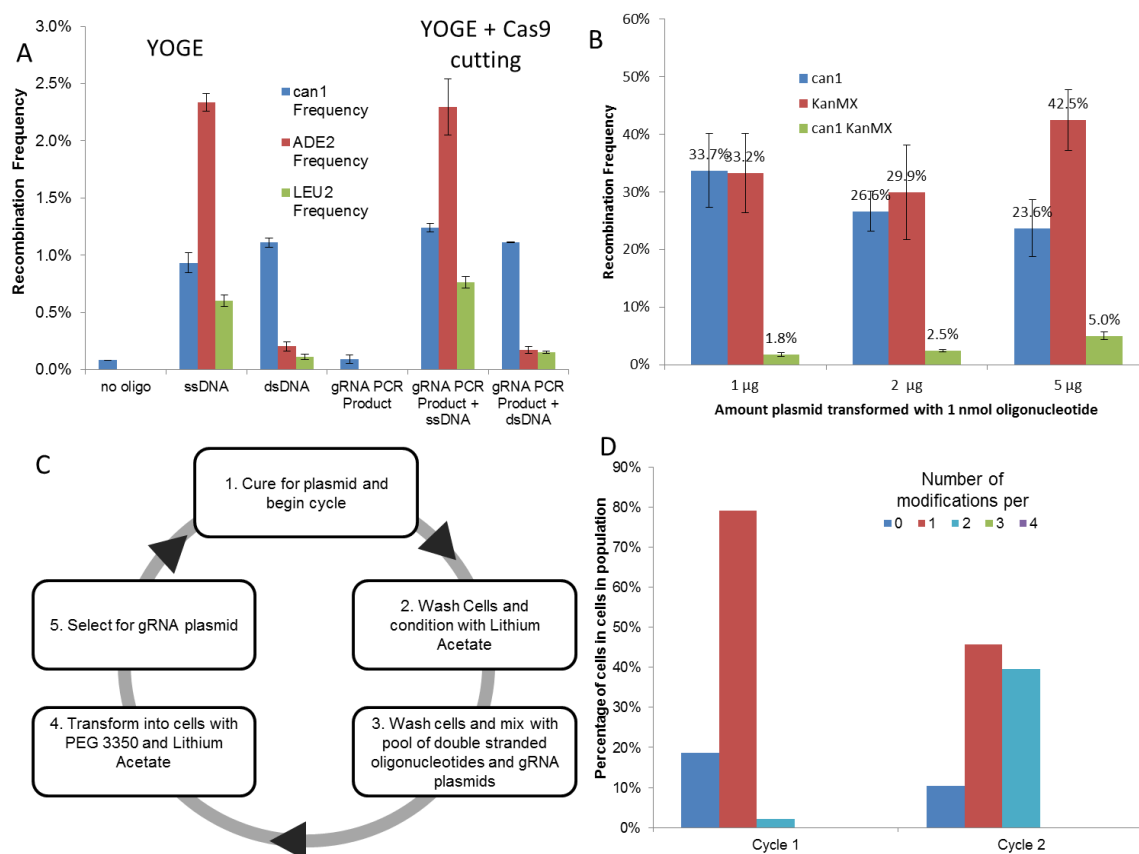


Figure 4.1 Optimization of Multiplex Genome engineering protocol A. Oligonucleotide recombination using YOG or YOG in combination with Cas9 cleavage directed by gRNA PCR products. B. Optimization of Cas9 oligonucleotide mediated genome engineering for two loci. C. Diagram of cycling procedure for Cas9 genome engineering in yeast. D. Percentage of recombinant cells in the population in of two cycles of Cas9 genome engineering while targeting four alleles important in thermotolerance.

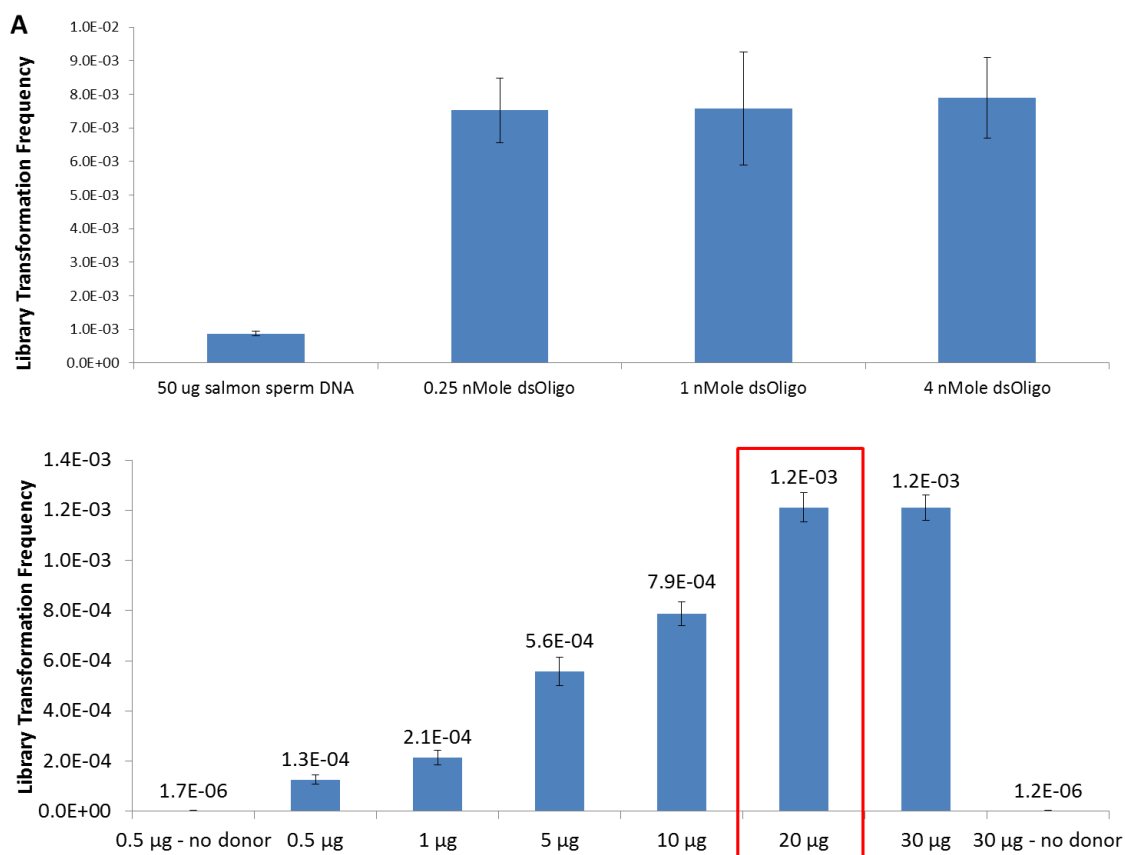


Figure 4.2 Optimization of gRNA/Cas9 plasmid and donor double stranded oligonucleotide delivery. A. Titration of double stranded oligonucleotides concentration using 20 ug of gRNA/Cas9 plasmid. B. Titration of gRNA/Cas9 plasmid using 0.25 nMole Oligonucleotide.

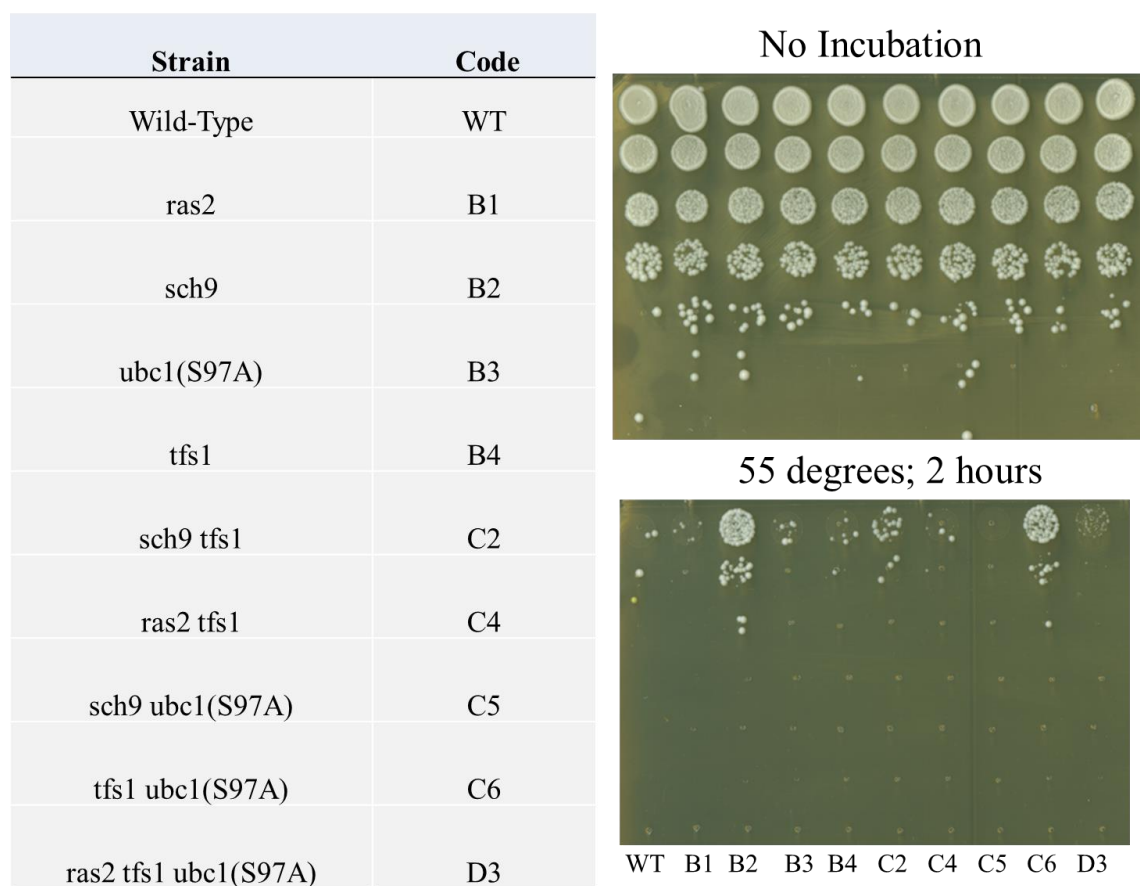


Figure 4.3 Thermotolerant mutant phenotypes and multiplex cassette insertion. A. Mutant strains isolated and ability to survive two hour heat shock at 55°C. Multiplexability of large cassette insertion using Cas9, with one selection marker on the gRNA plasmids

Chapter 5 – RNA-guided gene drives can efficiently and reversibly bias inheritance in *S. cerevisiae*

5.1 Introduction

Gene drives have the potential to address diverse ecological problems by altering the traits of wild populations. As 'selfish' genetic elements, they spread not by improving the reproductive fitness of the organism, but by increasing the odds that they themselves will be inherited. Because this inheritance advantage can counteract the fitness costs associated with the drive itself or with adjacent genes carried along with it, they are theoretically capable of 'driving' unrelated traits through

Inheritance-biasing is a common strategy in nature. (22) One elegant class of inheritance-biasing genes spreads by cutting homologous chromosomes that do not contain them, thereby inducing the cellular repair process to copy them onto the damaged chromosome by homologous recombination (**Figure 5.1A**). This process is known as 'homing'.(69) The best-known homing endonuclease gene is I-SceI, whose product cuts the gene encoding the large rRNA subunit of *S. cerevisiae* mitochondria. Most are capable of homing with extremely high efficiencies; I-SceI is correctly copied 99% of the time.(70) Austin Burt first suggested that homing endonucleases might be used to construct synthetic gene drives capable of altering wild populations of multicellular organisms in 2003 (**Figure 5.1B**)(23). Laboratories subsequently reported that the I-SceI endonuclease gene exhibited homing in transgenic laboratory populations of mosquitoes(71) or fruit flies (72, 73) with an I-SceI recognition site inserted into the corresponding locus. However, gene drives based on homing endonucleases are

constrained by the difficulty of retargeting these enzymes to cleave useful sequences within wild-type genomes.(74)

The recent development of the CRISPR nuclease Cas9, which cleaves target sequences specified by “guide RNA” molecules, has enabled scientists to edit the genomes of diverse species.(27, 44, 45, 56, 62, 75–78) Because Cas9 can apparently be used to edit any gene, the question of whether it can be used to gene drives capable of spreading those changes through wild populations is highly relevant. Our lab previously detailed the theoretical potential for RNA-guided gene drives to alter wild populations, including an evolutionary analysis, novel drive architectures, and containment measures robust to human error.(79) However, these new architectures and control strategies have not yet been validated. Indeed, whether Cas9 can bias inheritance at all remains unknown, raising the question of whether initiating public discussions and engaging in regulatory reform are immediately necessary.(80) We sought to address these issues by constructing several types of RNA-guided gene drives and quantifying their ability to bias inheritance in the yeast *S. cerevisiae*.

5.2 Molecularly Contained Gene Drives

Because gene drives have the capacity to alter native populations, we took stringent precautions to prevent the accidental escape of our gene drives into wild yeast. We first employed a method of molecular containment¹ in which we split our Cas9 based gene drive system into two physically separate genetic parts: an episomally encoded Cas9 gene and a drive element encoding the guide RNA. This allowed us to avoid creating a self-sufficient inheritance-biasing cassette while still targeting wild-type yeast strains.

Though simple, this form of molecular containment is not vulnerable to human error; even if drive-containing yeast were to escape into the wild, the required Cas9 episomal plasmid would rapidly be segregated away from the drive element, rendering the drive inoperative.

To directly measure the efficiency of Cas9 gene drives in yeast, we used the ADE2 gene as a visual marker.⁽⁸¹⁾ If red *ade2Δ* haploids are mated with cream-colored wild-type haploids, the resulting heterozygous diploids inherit one functional copy of ADE2 and are cream-colored. When these diploids undergo meiosis and reproduce via sporulation, half the resulting haploids inherit the broken copy and are consequently red; the other half inherit the intact copy and are cream-colored (**Figure 5.2A**). But if the red haploids encode a functional gene drive in place of ADE2, it will cut and replace the intact ADE2 locus inherited from the wild-type parent, yielding red diploids. Following meiosis, all haploid progeny will inherit one of the two gene drive alleles and will also be red (**Figure 5.2B**). Thus, the cutting efficiency of a gene drive that replaces ADE2 can be assessed by mating drive-containing haploids to wild-type haploids, selecting for diploids and counting the fraction that are red.

We built a Cas9-based gene drive targeting ADE2 by placing a guide RNA against the wild-type ADE2 gene in place of the endogenous ADE2 locus. We mated these red *ade2::sgRNA* haploids to wild-type yeast of the opposite mating type in the presence or absence of the Cas9 plasmid and plated on media that selects for diploids. Nearly all diploid colonies were red when the Cas9 plasmid was present, indicating highly efficient cutting of the ADE2 copy inherited from the wild-type parent (**Figure**

5.2C). As expected, we did not observe red diploid colonies in the absence of Cas9, demonstrating that the drive only functions in populations encoding both Cas9 and guide RNA.

To verify that the ADE2 alleles from drive-containing diploids were indeed lost, mated diploids were sporulated and their resultant haploid progeny were examined. Upon dissecting 18 cas9+ diploids, we observed a perfect 4:0 ratio of red:cream haploids, confirming that all copies of the ADE2 locus were disrupted. In contrast, 18 cream-colored cas9- diploids yielded a 2:2 red:cream ratio, indicating normal inheritance of the inactivated drive and the wild-type ADE2 allele (**Figure 5.2D**). To determine whether ADE2 disruptions in red diploids were the result of successful copying of the drive element, we sequenced the 72 haploids derived from dissected cas9+ diploids. All sequenced colonies contained intact drives without additional mutations, indicating that drive mobilization was efficient and occurred at high fidelity.

We next tested whether RNA-guided gene drives could be designed to bias the inheritance of not only the minimal drive element, but also any closely associated “cargo” gene whose spread through an existing population may be desirable. As a proof of principle, we inserted the URA3 gene in cis to the ade2::sgRNA drive element. URA3 allows laboratory modified yeast strains to grow in the absence of uracil supplementation (**Figure 5.3A**). We mated these URA3-containing drive haploids to wild-type haploids in the presence of an episomal Cas9 plasmid, selected diploids (all of which were red), sporulated them, and dissected 18 tetrads. As was the case for the original ADE2 gene

drive, all of the sporulated haploid cells formed red colonies. Crucially, all grew normally when replica plated on uracil deficient media, indicating that URA3 was efficiently copied with the drive (**Figure 5.3B**).

We subsequently sought to determine whether gene drives cannot simply disrupt a gene, but recode it and leave the function intact. A gene drive targeting an essential gene in this manner should be more evolutionarily stable than one that cuts a non-essential gene such as ADE2, which can be readily blocked by mutations that mutate or remove the target site¹. We consequently built a drive targeting the essential ABD1 gene (**Figure 5.3C**). (82) We mated this haploid strain, which has a recoded ABD1 allele upstream of the guide RNA, to wildtype cells in the presence of Cas9. We then selected diploids, sporulated the cells, dissected 18 of them, and sequenced the 72 resulting segregants. All contained the drive element and recoded ABD1 locus, thereby validating our proposed essential gene recoding architecture.

5.3 Gene Drives Spread into Phylogenetically Diverse Yeast Strains

All of our earlier experiments involved matings between haploids of the same strain. We were curious whether our gene drives could be copied into a diverse group of wild *S. cerevisiae* strains at equal efficiency. These strains may vary in some of the many factors determining gene drive copying efficiency, such the types of repair machinery available to the cell at the time of the cut, the chromatin state of the locus, and the degree of homology flanking the double-strand break generated by the drive. We correspondingly mated ADE2 drive-containing haploids with 6 phylogenetically and phenotypically diverse wild-type strains of haploid *S. cerevisiae*. (83)

To more accurately measure the efficiency with which all of these gene drives were copied in various backgrounds, we performed quantitative PCR on populations of the resulting diploids using one set of primers specific to the drive and another set designed to amplify either wild-type alleles or those disrupted by non-homologous end-joining. The mean fraction of diploid chromosomes containing the ADE2 gene drive was over 99% regardless of wild-type parent strain (**Figure 5.4**), attesting to the robustness of the drive in diverse backgrounds. Of note, addition of the URA3 cargo gene did not appreciably change this efficiency. The drive that targets and recodes the essential ABD1 gene was similarly efficacious.

We next determined whether RNA-guided gene drives could bias inheritance over successive generations. Because *S. cerevisiae* reproduces mostly through asexual division and only a subset of the population sporulate in the laboratory even when induced to undergo meiosis, experiments to determine the long-term population-level efficiency of gene drives are difficult to perform and interpret.⁽⁸³⁾ As a surrogate for population takeover experiments, we sought to measure the performance of successive copying events by a single drive element. Towards this goal, we mated the haploid progeny of the ADE2 gene drive to wild-type haploids, selected for diploids, and ran quantitative PCRs. The constructs biased inheritance at the same efficiency in the second generation as they did in the first (**Figure 5.4**, far right).

5.4 Complete Gene Drives and Reversibility

We designed the previously described gene drives to be incapable of autonomous spread due to their requirement for exogenously supplied Cas9 (**Figure 5.2C-D**, **Figure**

5.4). To determine whether an autonomous drive encoding the large cas9 gene is similarly efficient, we constructed one such drive targeting a recoded synthetic sequence within ADE2 (**Figure 5.5A**). We further sought to determine whether the loss of ADE2 function induced by this drive might be undone using a “reversal drive”¹. We consequently built a second autonomous gene drive to cleave the first autonomous ADE2-disrupting drive and subsequently restore an intact copy of the ADE2 gene (**Figure 5.5B**). Quantitative PCR demonstrated that both of these autonomous drives were copied at an average efficiency over 99% (**Figure 5.5D**). Together, these results demonstrate the efficacy and reversibility of autonomous Cas9- based gene drives.

5.5 Discussion and Impact on the Gene Drive Community

Our discovery that Cas9 can bias inheritance in diverse wild yeast strains over successive generations at very high efficiency demonstrates that RNA-guided gene drives can function in eukaryotic organisms as predicted. By itself this does not guarantee the success of gene drives in other organisms as the rate of homologous recombination varies between species and is known to be particularly high in yeast. However, the fact that we observed inheritance biasing rates equal to or exceeding that of the natural I-SceI homing endonuclease gene is highly promising. Because a drive based on I-SceI efficiently initiated copying following successful genome cutting events in transgenic *Anopheles gambiae* mosquitoes that contained an I-SceI recognition site⁶ and Cas9 typically cuts more efficiently than does I-SceI, our findings suggest that RNA-guided gene drives will be highly effective at suppressing or spreading antimalarial alleles through populations of this important malaria vector. (23, 24, 71, 72, 84–87). Similarly, RNA-guided gene

drives are likely to be more effective than those based on I-SceI, zinc-finger nucleases, or TALENs due to their ability to efficiently cut multiple sequences without requiring highly repetitive elements. (49, 50, 67)

The success of our ABD1 gene drive demonstrates the feasibility of targeting and recoding genes important for fitness, a strategy expected to improve the evolutionary stability of gene drives¹. We recommend that future efforts seeking to build gene drives intended for eventual release adopt this same approach. On a similar note, our successful test of a reversal drive construct strongly suggests that such drives should be constructed in tandem with any gene drive capable of enacting a specific change in a wild population as a safeguard against accidental escape.

More generally, our findings suggest that yeast may prove a useful platform for swiftly testing RNA-guided gene drive architectures before moving them into multicellular organisms. The power of yeast genetics and the ease of genome manipulation will facilitate combinatorial investigations into gene drive optimization. For example, studies might explore how biasing repair pathway choice affects the efficiency of copying for gene drives of various sizes. Because the factors involved in these pathways are broadly conserved, these experiments could guide gene drive optimization in other organisms. (88, 89)

While highly encouraging for potential gene drive applications, our results also sound a note of caution for subsequent experiments. That our drives were readily copied into a variety of yeast strains collected from all over the world underscores the potential for a single gene drive to affect very large populations. Poor flanking homology is not an

effective barrier, at least in *S. cerevisiae*. Moreover, the ADE2 gene drive took only two weeks to design, build, and test, suggesting that many laboratories are capable of building gene drives in yeast. Since yeast reproduce mainly through asexual division, gene drives would need to be considerably less costly to organismal fitness in order to spread in the wild than would a comparably efficient gene drive in an organism that always reproduces via mating. However, natural endonuclease gene drives such as I-SceI do exist within yeast. Whether our gene drives or the typical RNA-guided gene drive will constitute this level of burden is as yet unknown.

It is more difficult to edit the genomes of multicellular model organisms such as *Drosophila* than it is for yeast, and still more difficult to alter those species for which gene drive applications are most likely to be relevant. However, a growing number of laboratories now make heritable alterations in more than a dozen sexually reproducing species. This confluence of factors demands caution. Because synthetic gene drives would alter the global environmental commons, the decision to deploy such a drive must be made collectively by society. Any accidental release could severely damage public trust in scientists. As demonstrated by numerous containment breaches involving pathogenic viruses and bacteria, physical methods of containment are always susceptible to human error and should not be exclusively relied upon whenever alternatives are available.

All scientists making heritable alterations with Cas9 should therefore employ non-physical containment methods sufficient to prevent the creation of an RNA-guided gene drive capable of spreading in the wild. Even scientists not intending to work with gene

drives should consider taking precautions, since any unintended insertion of the Cas9 gene and guide RNAs near a targeted site could generate a gene drive. Fortunately, a simple and costless precaution is both available and already utilized for different reasons by many laboratories: avoid delivering the Cas9 gene on a DNA cassette that also encodes a guide RNA. As we have shown, guide RNAs alone cannot bias inheritance in the absence of Cas9 and consequently cannot spread through wild populations (**Figure 5.4B**).

In conclusion, our demonstration of diverse gene drive architectures enabling Cas9-mediated inheritance biasing in wild *S. cerevisiae* can guide efforts to build RNA-guided gene drives in other organisms and underscores the urgent need for precautionary control strategies, inclusive public engagement, and regulatory reform in advance of real-world applications.(80) Additionally, the use of gene drives in many yeast species could be useful for generation of knockouts, or mutants of genetically intractable strains. When two gene drives strains are mated the gene drives should propagate into the new chromosome, and created double gene drive offspring. These combinations of gene drives could create a combinatorial gene drive library that could be useful for studying epistasis or in directed strain design

5.6 Materials and Methods

5.6.1 Physical Confinement of Gene Drives

All experiments were carried out using a strict barrier protocol. Plates and tubes containing gene drive strains were kept in separate areas of bench space, cold room space, and freezer space. All plates were securely parafilmed and bench space cleaned

daily using a 70% ethanol solution. Exposed glassware was first soaked for 24 hours in Wescodyne solution before cleaning. All waste was bagged in biohazard specific bags, sealed, and incinerated.

5.6.2 Plasmids and genomic cassettes

Gene drive cassettes were synthesized from gBlocks (Integrated DNA Technologies, Coralville, IA) and inserted into SK1 cells via Cas9-mediated genome modification as follows. Guide RNAs for each drive were cloned into p416-Cas9 containing plasmids with expression driven by the SNR52 promoter¹⁸. 60 base pair homology arms to the target locus were added on both ends of the gene drive cassette via PCR and 5 ug of PCR product was co-transformed with the p416-Cas9-gRNA plasmids. Correctly integrated gene drives were verified by sequencing and p416-Cas9-gRNA plasmids were cured using 5-Fluoroorotic Acid (FOA) selection.

To create the *URA3*-containing *ADE2* gene drive, the *ADE2* gene drive was cloned next to the *Candida albicans* *URA3* gene in the pAG60 plasmid. The entire *URA3* cassette and gene drive were PCR amplified and inserted using Cas9-mediated genome modification into the *ADE2* locus of haploid SK1 cells.

The recoded C-terminus of the *ABDI* gene and corresponding gene drive were synthesized as a gBlock to remove homology and generate mutations in the seed sequence via synonymous changes. The TEF1 terminator was inserted at the 3' end of the recoded *ABDI* gene between the gene and the gRNA as *ABDI* shares a terminator with the *VHCl* gene. The entire cassette was integrated into the haploid SK1 genome using Cas9-mediated genome modification.

The ADE2 gene was recoded by cotransforming a double stranded oligonucleotide and a p416 plasmid containing Cas9 and a gRNA targeting the ADE2 region to recode. The oligonucleotide silently recoded the ADE2 gene and included an orthogonal target and PAM sequence. The complete gene drive (Cas9 and gRNA, targeting the recoded ADE2 gene) was generated by cloning a gRNA into the p416-Cas9 plasmid. An orthogonal genomic target was also included in the complete gene drive to later be targeted by the reversal drive. The Cas9 and gRNA linear construct was amplified by PCR using the same homology arms as the sole gRNA gene drive construct. The construct was co-transformed into By4723 cells with the plasmid it was amplified from and the cells were selected for uracil prototrophy. Correct integrations were screened via colony PCR. This plasmid was later removed using FOA.

The reversal drive (Cas9 and gRNA integrated upstream of the ADE2 gene) was generated by cloning an alternately encoded gRNA into a p414 plasmid containing Cas9. This alternatively encoded gRNA contains less homology to previously used gRNAs and would reduce the chance of unwanted recombination when used to replace the complete gene drive. This gRNA targets a 20bp region inserted with the complete gene drive. The TRP1 gene with the Cas9 and gRNA were PCR amplified with homology arms to the 5' region of the ADE2 and the product was transformed into SK1 A cells. Cells were selected for tryptophan prototrophy and screened via PCR for correct integrations.

The p416-Cas9-gRNA plasmid (conferring uracil prototrophy) is a variant of the previously described p414-Cas9-gRNA plasmid (conferring tryptophan prototrophy)¹⁸ (Addgene #43802). One or the other was used in each mating experiment. The pRS413

vector was transformed into select cell types to confer histidine prototrophy as a marker to select for diploid cells.

5.6.3 Sporulation and tetrad dissection

After mating in liquid YPAD and selection for diploids on selection plates, the selection plates were scraped into 10 mL selective media and grown overnight at 30°C. A fresh 5 mL YPAD culture was then inoculated to and OD=0.1 and grown 4-5 hours at 30°C. The entire culture was then washed twice in 10 mL water, inoculated into 2 mL of sporulation media (1% potassium acetate), and incubated at room temperature for 3 days or until spores were visible. Sporulated cells were suspended in 50 µL of a stock solution of zymolyase (50 µg/mL in 1M sorbitol) and incubated at 30C for 5 minutes, transferred to ice, diluted with 150 µL cold H₂O, microdissected using a Zeiss tetrad dissection microscope, and isolated spores grown on YPAD plates. **Table A1** shows the matings and the selection used for diploids.

5.6.4 Selection for URA3 function

Dissected spores were grown in synthetic complete (SC) media and then spotted onto SC medium as well as SC medium without uracil. To enhance red color, all SC solid media used for plate images contained 0.5 X adenine hemisulfate (final concentration of 0.08 mM).

5.6.5 Verification of chromosomal segregation

Three genes on chromosome 15 flanking the ADE2 gene were sequenced in two dissected tetrads of the complete gene drive cross. VAM3, TRS33, and DPP1 were

amplified using PCR from colonies and sequenced using Sanger sequencing.

5.6.6 Quantitative PCR

Candidate primer pairs were designed to amplify short regions specific to each drive or the wild-type sequence replaced by the drive, as well as the *ACT1* gene as a control. Genomic DNA was extracted using Method A as described in Looke et al. (90) KAPA SYBR FAST qPCR Master Mix (2X) was used to perform the qPCR reaction along with 25 ng of genomic DNA. The amplification efficiency and relative specificity of each primer pair were measured by amplifying dilutions of genomic DNA from wild-type and drive haploids, respectively, and the best-performing and well-matched pairs selected for use (see below for all primers used). Quantitative PCR reactions were performed on genomic DNA isolated from each parental haploid as well as from diploids arising from three independent mating events. Three reactions (technical replicates) were performed per sample on a LightCycler 96 machine by Roche.

5.6.7 qPCR Calculations

Results from three technical replicates were averaged for calculations. In order to directly calculate the ratio of alleles before PCR amplification, we first determined the efficiencies of the different primer pairs. Efficiencies were calculated from qPCR runs of serial dilutions (6 orders of magnitude) as: $\text{Efficiency} = 10^{-1/\text{slope}}$ R^2 values were higher than 0.99 in all cases except for one pair (*ade2::URA3+sgRNA*).

The allelic ratios were calculated as:

$$x_a \cdot E_a^{\text{Ct},a} = x_b \cdot E_b^{\text{Ct},b}$$

$$x_a / x_b = E_b^{C_{t,b}} / E_a^{C_{t,a}}$$

with x_a and x_b being the initial concentration of drive and Wild-type DNA,

E_a and E_b the efficiency of the respective primer pairs and

$C_{t,a}$ and $C_{t,b}$ the C_t values for each sample.

Figure 4B was generated using BoxPlot. (91)

5.7 Figures

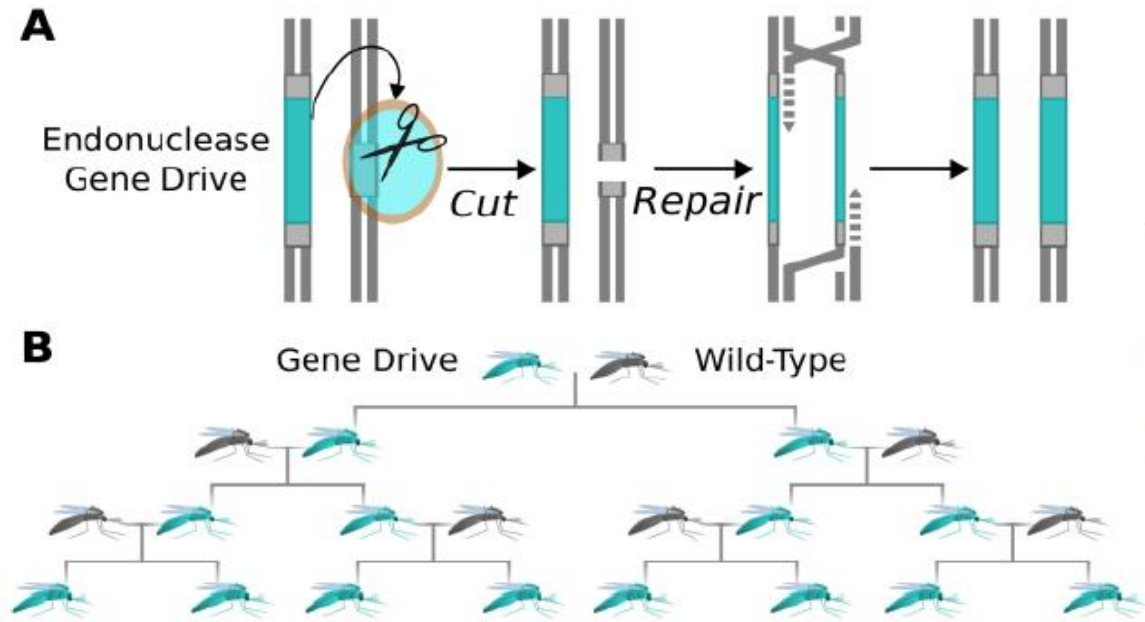


Figure 5.1 Mechanism and population-level effect of endonuclease gene drives. (A) Homing endonucleases cut competing alleles, inducing the cell to repair the damage by copying the endonuclease gene. (B) By converting heterozygous germline cells into homozygotes containing two copies (teal), gene drives increase the odds that they will be inherited and consequently spread themselves and associated changes through wild populations (grey).

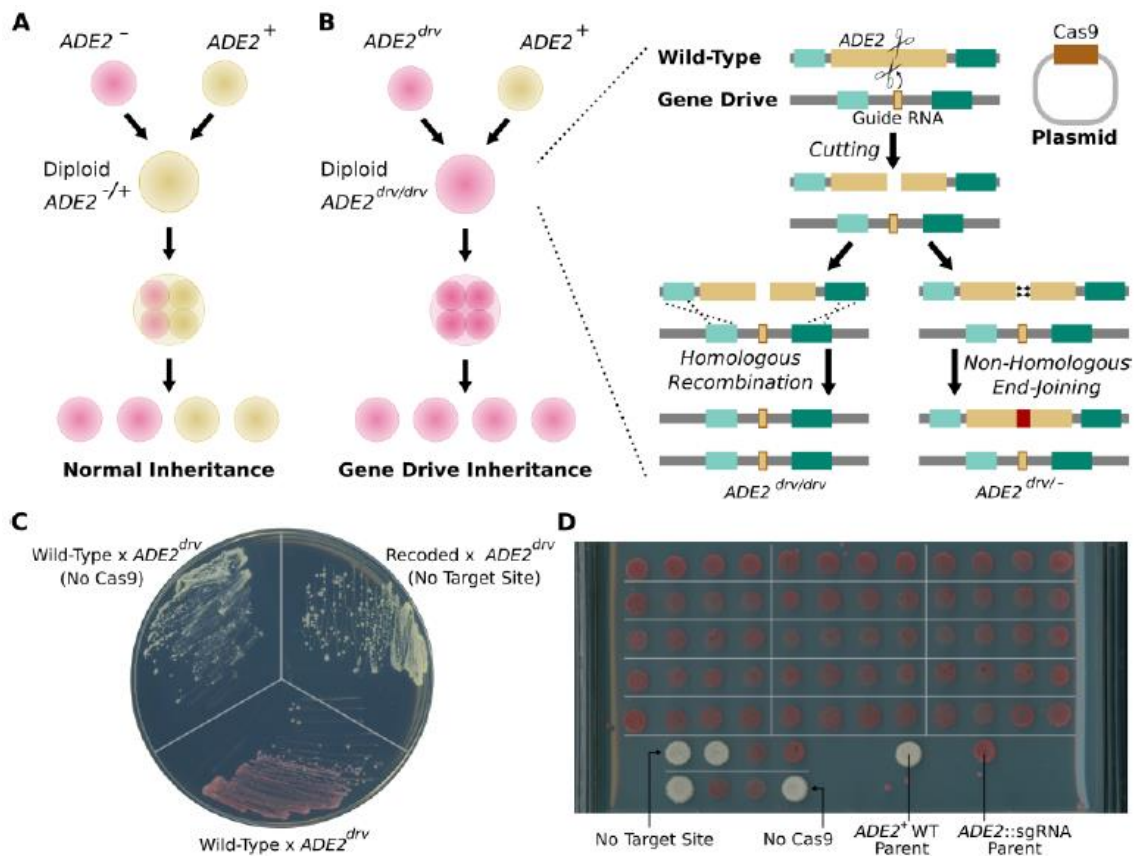


Figure 5.2 Biased inheritance of *ADE2* is readily visible in *S. cerevisiae*. (A) Mutations in *ADE2* generate a red phenotype on adenine-limiting media due to the buildup of red pigments. Mating a red mutant haploid to a wild-type haploid produces cream-colored diploids, which yield 50% red and 50% cream-colored progeny upon sporulation. (B) When haploids with a gene drive targeting *ADE2* mate with wild-type haploids in the presence of Cas9, cutting and subsequent replacement or disruption of *ADE2* produces red diploids that upon meiosis yield exclusively red progeny. (C) Diploids produced by mating wild-type and *ade2::sgRNA* gene drive haploids yield cream-colored colonies in the absence of Cas9 or when the target site is removed by recoding but uniformly red colonies when both are present, demonstrating Cas9-dependent disruption of the wild-type *ADE2* copy. (D) Spores from 15 dissected tetrads produce uniformly red colonies on adenine-limited plates, confirming disruption of the *ADE2* gene inherited from the wildtype parent. In the absence of the target site or Cas9, normal 2:2 segregation is observed.

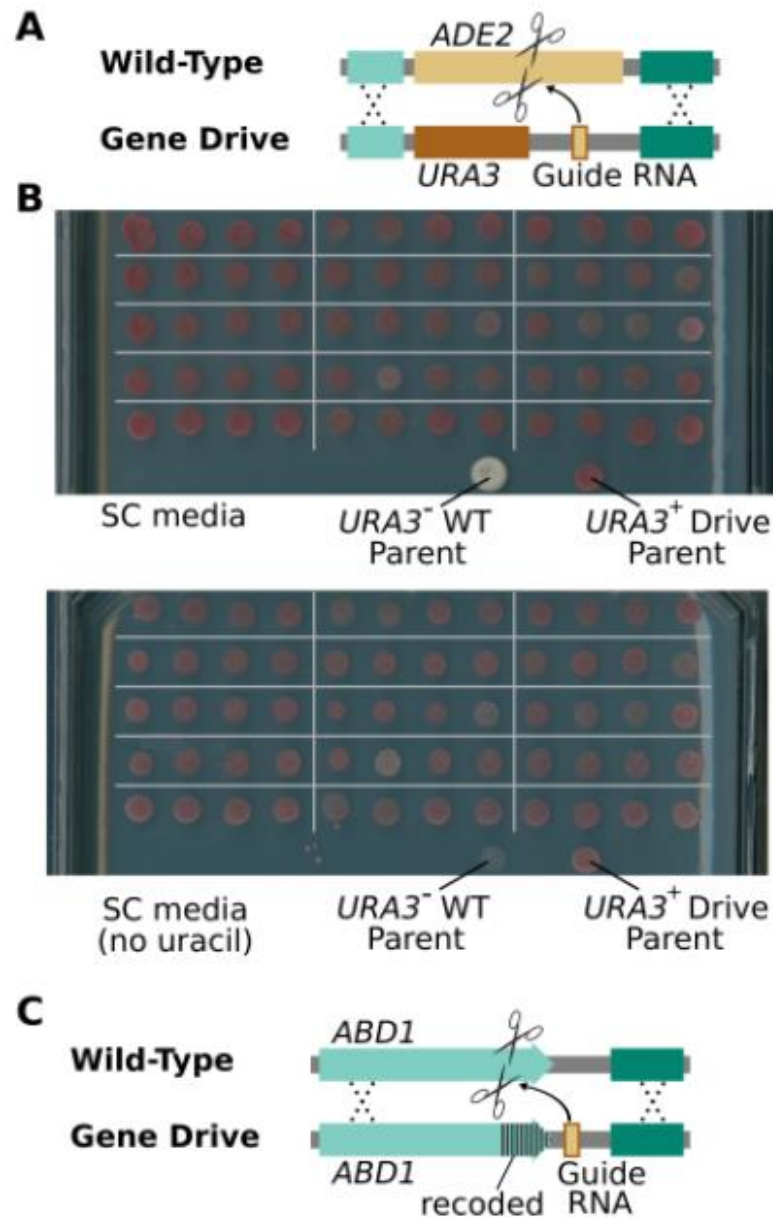


Figure 5.3 Gene drives and cargo genes remain intact upon copying and can spread by targeting both non-essential and essential genes. (A) The *ADE2*-targeting gene drive was modified to carry *URA3* as a cargo gene. (B) Diploids produced by mating wild-type *URA3*⁻ haploid yeast with haploids encoding the gene drive carrying *URA3* were sporulated and tetrads dissected to isolate colonies arising from individual spores. Pictures are spores from 15 of these tetrads. All grew when replica-plated onto plates lacking uracil, demonstrating that the drive successfully copied *URA3* in all diploids. (C) Depiction of a gene drive designed to cut and recode the 3' end of the essential *ABD1* gene.

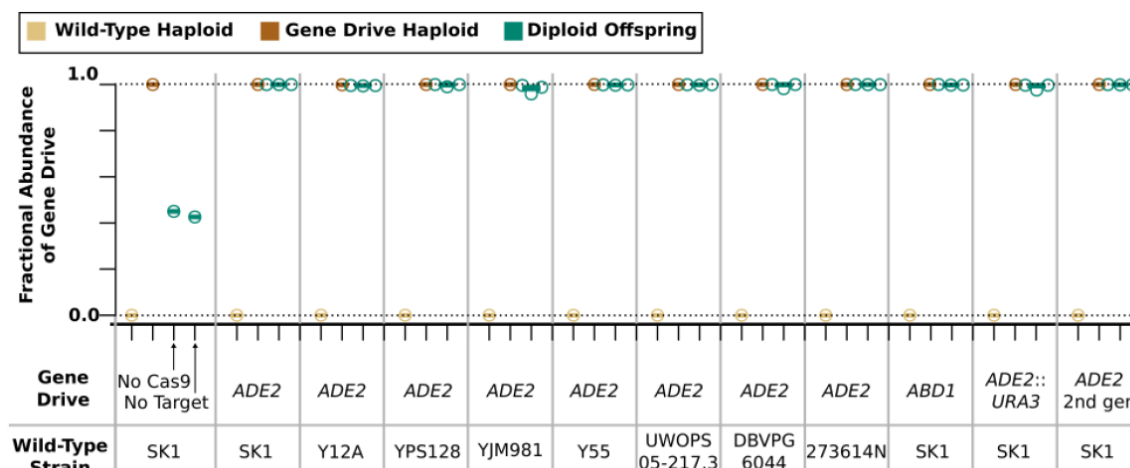


Figure 5.4 qPCR data for yeast gene drives. The extent of inheritance-biasing induced by various gene drives in diverse yeast strains as measured by quantitative PCR. Results depict the relative abundance of wild-type and drive-containing alleles in diploids arising from matings between SK1 haploids bearing gene drives and diverse wild-type haploid strains. “No Cas9” and “No Target” refer to haploid cells containing the ADE2 drive mated to wildtype haploids in the absence of Cas9 or to an otherwise wild-type strain with Cas9 that has a mutation in the targeted sequence that blocks cutting. “2nd gen” refers to the haploid progeny of an earlier mating.

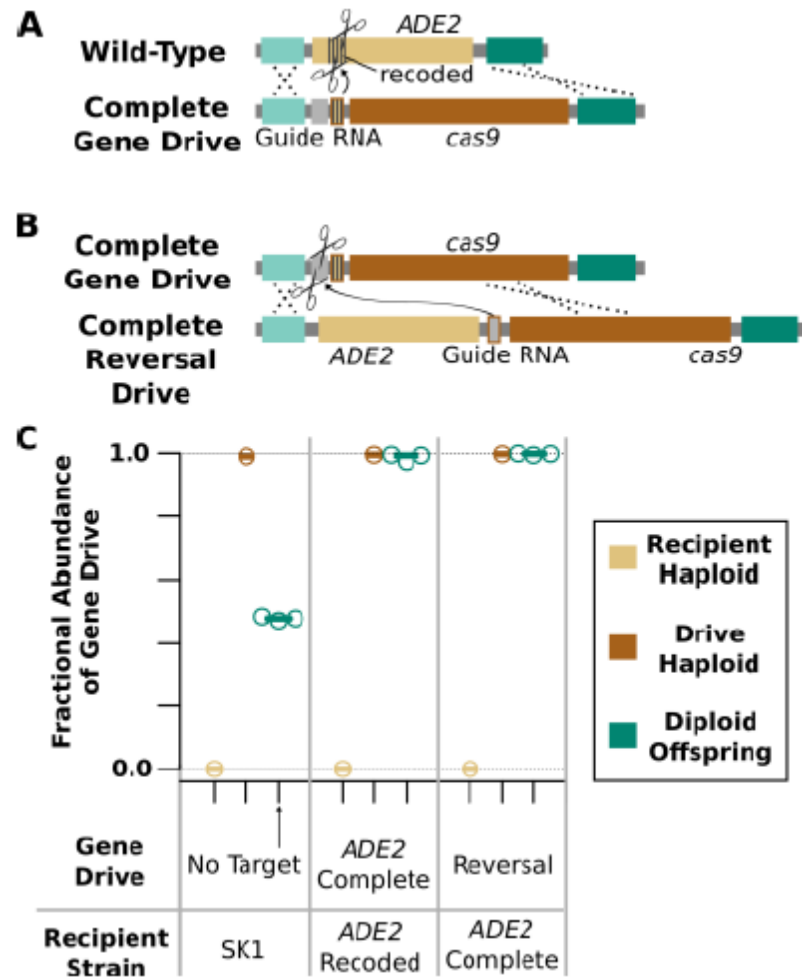


Figure 5.5 Essential, autonomous, and reversal gene drives. (A) A complete and autonomous gene drive that cuts and replaces the recoded ADE2 gene. (B) A reversal drive that cuts the autonomous drive and restores ADE2. (C) Quantitative PCR results depicting the relative abundance of wild-type and drive-containing alleles in diploids arising from matings between SK1 haploids bearing the above gene drives and wild-type SK1 yeast.

CHAPTER 6 - Discussion

Modification of yeast genomes has been important classically to study eukaryotic functional genomics and production of industrial compounds. In recent years, more and more labs are using yeast as an assembly tool for genomes of other species. (9, 19, 90) This work has aimed to develop methods for yeast genome engineering that would allow for rapid and simple modification of genomes in living yeast cells. While YOG and Cas9-mediated genome engineering, both generate library sizes of 10^5 individuals, each has its own strength for different applications. For diversity generation, YOG may be more appropriate as the cycling between transformations is relatively fast (~12 hours). For precision genome engineering with low background, the Cas9-mediated approach can generate a population that is near isogenic for the desired changes.

This dissertation presents work that was the first to use CRISPR/Cas9 systems in budding yeast. Furthermore, the demonstration of RNA-guided gene drives in this dissertation also represents the first released experiments showing that Cas9 based gene drives function in eukaryotic organisms. Since testing both CRISPR/Cas9 and gene drives in yeast, several groups have gone to use these tools in other organisms. (62, 63, 75, 92, 93) Gene drives could have many uses for pushing alleles through wild populations, as well as for genome engineering in yeast for academic and industrial uses. If a combinatorial library is desired with many members, using multiple gene drives may be useful as several rounds of mating could result in large combinatorial libraries (depending on number of combinations) with high frequency. Furthermore, future work using RNA-guided gene drives in yeast could employ a knockout library comprised of

gene drives to generate minimal yeast genomes. Using Cas9 for prescient and high efficiency genome engineering has been shown in even recalcitrant yeasts.(61, 92) It is likely that future projects will employ Cas9 to generate libraries of mutant strains for functional genomics and industrial applications.

While the technologies I describe in this dissertation were tested in yeast, their applicability is general. The development of enable tools for synthetic biology as it allows more questions to be answered and more useful organisms to be constructed. As we push the boundaries of designing and synthesizing life, we can begin to imagine a world in which biological design is not constrained by current dogma and can transcend our modern utilization and understanding of living systems.

APPENDIX

MATa Genotype	MATα Genotype	Selection
SK1 pRS414 – Cas9	SK1 ade2::gRNA (gene drive), pRS413	SC-histidine - tryptophan
SK1 ade2::gRNA + URA3 (gene drive), p414-Cas9	SK1 pRS413	SC-histidine - tryptophan
SK1 p414-Cas9	SK1 abd1::ABD1 recoded +gRNA (gene drive), pRS413	SC-histidine - tryptophan
Y12A Hygromycin B resistance (HygR)	SK1 ade2::gRNA (gene drive), p416-Cas9	SC-uracil+300 ug/mL Hygromycin B
YPS128 Hygromycin B resistance (HygR)	SK1 ade2::gRNA (gene drive), p416-Cas9	SC-uracil+300 ug/mL Hygromycin B
YJM981 Hygromycin B resistance (HygR)	SK1 ade2::gRNA (gene drive), p416-Cas9	SC-uracil+300 ug/mL Hygromycin B
Y55 Hygromycin B resistance (HygR)	SK1 ade2::gRNA (gene drive), p416-Cas9	SC-uracil+300 ug/mL Hygromycin B
UWOPS05-217.3 Hygromycin B resistance (HygR)	SK1 ade2::gRNA (gene drive), p416-Cas9	SC-uracil+300 ug/mL Hygromycin B
DBVPG 6044 Hygromycin B resistance (HygR)	SK1 ade2::gRNA (gene drive), p416-Cas9	SC-uracil+300 ug/mL Hygromycin B
273614N Hygromycin B resistance (HygR)	SK1 ade2::gRNA (gene drive), p416-Cas9	SC-uracil+300 ug/mL Hygromycin B
SK1 ADE2:: ADE2 silently recoded genomic target seed sequence, p414-Cas9	SK1 ade2::gRNA (gene drive), pRS413	SC-histidine - tryptophan
SK1, p414-empty	SK1 ade2::gRNA (gene drive), pRS413	SC-histidine - tryptophan
BY4723 ADE2 recoded	SK1 ade2::complete gene drive (Cas9+gRNA,with recoded ADE2 target)	SC-histidine - tryptophan
BY4723	SK1 ade2::complete gene drive (Cas9+gRNA,with recoded ADE2 target)	SC-histidine - tryptophan
Complete Gene drive Haploid from dissected tetrad, p413-empty	SK1 ADE2 +reversal gene drive (Cas9+gRNA, targeting inserted orthogonal target), p416-empty	SC-histidine -uracil

Table A1. List of matings for yeast gene drive experiments.

BIBLIOGRAPHY

1. Danna K, Nathans D (1971) Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proceedings of the National Academy of Sciences of the United States of America* 68(12):2913–2917.
2. Van Erp PB, Bloomer G, Wilkinson R, Wiedenheft B (2015) The history and market impact of CRISPR RNA-guided nucleases. *Current Opinion in Virology* 12:85–90.
3. Esvelt KM, Wang HH (2013) Genome - scale engineering for systems and synthetic biology. *Molecular Systems Biology* 9(1):641.
4. Agapakis CM (2014) Designing synthetic biology. *ACS Synthetic Biology* 3(3):121–128.
5. Julleson D, David F, Pflieger B, Nielsen J (2015) Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnology Advances*.
6. Keung AJ, Joung JK, Khalil AS, Collins JJ (2015) Chromatin regulation at the frontier of synthetic biology. *Nature Reviews. Genetics* 16(3):159–171.
7. Samuel D (1993) Ancient Egyptian Cereal Processing: Beyond the Artistic Record. *Cambridge Archaeological Journal* 3(2):276–283.
8. Samuel (1996) Investigation of Ancient Egyptian Baking and Brewing Methods by Correlative Microscopy. *Science* 273(5274):488–490.
9. Muller H, et al. (2012) Assembling large DNA segments in yeast. *Methods in Molecular Biology* 852:133–150.
10. Barnett JA (2007) A history of research on yeasts 10: foundations of yeast genetics1. *Yeast* 24(10):799–845.
11. Müller B, Grossniklaus U (2010) Model organisms — A historical perspective. *Journal of Proteomics* 73(11):2054–2063.
12. Hinnen A, Hicks JB, Fink GR (1978) Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America* 75(4):1929–1933.
13. Wach A, Brachat A, Pöhlmann R, Philippsen P (1994) New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* 10(13):1793–1808.

14. Storici F, Durham CL, Gordenin DA, Resnick MA (2003) Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 100(25):14994–14999.
15. Moerschell RP, Tsunasawa S, Sherman F (1988) Transformation of yeast with synthetic oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* 85(2):524–528.
16. Brachman EE, Kmiec EB (2003) Targeted Nucleotide Repair of *cyc1* Mutations in *Saccharomyces cerevisiae* Directed by Modified Single-Stranded DNA Oligonucleotides. *Genetics* 163(2):527–538.
17. Liu L, Cheng S, van Brabant AJ, Kmiec EB (2002) Rad51p and Rad54p, but not Rad52p, elevate gene repair in *Saccharomyces cerevisiae* directed by modified single-stranded oligonucleotide vectors. *Nucleic Acids Research* 30(13):2742–2750.
18. Gibson DG (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Research* 37(20):6984–6990.
19. Gibson DG, et al. (2008) 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 of the United States of America* 105(51):20404–20409.
20. Wang HH, et al. (2012) Genome-scale promoter engineering by coselection MAGE. *Nature Methods* 9(6):591–593.
21. Wang HH, et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460(7257):894–898.
22. Genes in Conflict — Austin Burt, Robert Trivers | Harvard University Press
Available at: <http://www.hup.harvard.edu/catalog.php?isbn=9780674027220>
[Accessed May 8, 2015].
23. Burt A (2003) Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proceedings of the Royal Society of London. Series B, Biological Sciences* 270(1518):921–928.
24. Deredec A, Burt A, Godfray HCJ (2008) The Population Genetics of Using Homing Endonuclease Genes in Vector and Pest Management. *Genetics* 179(4):2013–2026.
25. Jinek M, et al. (2012) A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337(6096):816–821.

26. DiCarlo JE, et al. (2013) Yeast Oligo-Mediated Genome Engineering (YOGE). *ACS Synthetic Biology* 2(12):741–749.
27. DiCarlo JE, et al. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research* 41(7):4336–4343.
28. DiCarlo JE, Chavez A, Dietz SL, Esvelt KM, Church GM (2015) RNA-guided gene drives can efficiently and reversibly bias inheritance in wild yeast. *bioRxiv*:013896.
29. Pijkeren J-P van, Britton RA (2012) High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Research*. Available at: <http://nar.oxfordjournals.org/content/early/2012/02/09/nar.gks147> [Accessed June 28, 2013].
30. Larionov V, Kouprina N, Solomon G, Barrett JC, Resnick MA (1997) Direct isolation of human BRCA2 gene by transformation-associated recombination in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 94(14):7384–7387.
31. Larionov V, Kouprina N, Nikolaishvili N, Resnick MA (1994) Recombination during transformation as a source of chimeric mammalian artificial chromosomes in yeast (YACs). *Nucleic Acids Research* 22(20):4154–4162.
32. Nijkamp JF, et al. (2012) De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microbial Cell Factories* 11(1):36.
33. Kow YW, Bao G, Reeves JW, Jinks-Robertson S, Crouse GF (2007) Oligonucleotide transformation of yeast reveals mismatch repair complexes to be differentially active on DNA replication strands. *Proceedings of the National Academy of Sciences of the United States of America* 104(27):11352–11357.
34. Hombauer H, Campbell CS, Smith CE, Desai A, Kolodner RD (2011) Visualization of Eukaryotic DNA Mismatch Repair Reveals Distinct Recognition and Repair Intermediates. *Cell* 147(5):1040–1053.
35. Wang HH, Xu G, Vonner AJ, Church G (2011) Modified bases enable high-efficiency oligonucleotide-mediated allelic replacement via mismatch repair evasion. *Nucleic Acids Research*. Available at: <http://nar.oxfordjournals.org/content/early/2011/05/23/nar.gkr183> [Accessed June 28, 2013].
36. Costantino N, Court DL (2003) Enhanced levels of λ Red-mediated recombinants in mismatch repair mutants. *Proceedings of the National Academy of Sciences of the United States of America* 100(26):15748–15753.

37. Liu L, Maguire KK, Kmiec EB (2004) Genetic re-engineering of *Saccharomyces cerevisiae* RAD51 leads to a significant increase in the frequency of gene repair in vivo. *Nucleic Acids Research* 32(7):2093–2101.
38. Zhang X-P, Galkin VE, Yu X, Egelman EH, Heyer W-D (2009) Loop 2 in *Saccharomyces cerevisiae* Rad51 protein regulates filament formation and ATPase activity. *Nucleic Acids Research* 37(1):158–171.
39. Rodriguez GP, Song JB, Crouse GF (2012) Transformation with Oligonucleotides Creating Clustered Changes in the Yeast Genome. *PLoS One* 7(8):e42905.
40. Liu L, et al. (2002) Strand Bias in Targeted Gene Repair Is Influenced by Transcriptional Activity. *Molecular and Cellular Biology* 22(11):3852–3863.
41. Wang HH, et al. (2012) Multiplexed in Vivo His-Tagging of Enzyme Pathways for in Vitro Single-Pot Multienzyme Catalysis. *ACS Synthetic Biology* 1(2):43–52.
42. Carroll D (2012) A CRISPR Approach to Gene Targeting. *Molecular Therapy* 20(9):1658–1660.
43. Bhaya D, Davison M, Barrangou R (2011) CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annual Review of Genetics* 45(1):273–297.
44. Mali P, et al. (2013) RNA-Guided Human Genome Engineering via Cas9. *Science*. Available at: <http://www.sciencemag.org/content/early/2013/01/03/science.1232033> [Accessed January 6, 2013].
45. Cong L, et al. (2013) Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*. Available at: <http://www.sciencemag.org/content/early/2013/01/03/science.1231143> [Accessed January 6, 2013].
46. Whelan WL, Gocke E, Manney TR (1979) The CAN1 Locus of *Saccharomyces cerevisiae*: Fine-Structure Analysis and Forward Mutation Rates. *Genetics* 91(1):35–51.
47. Bennett CB, Westmoreland TJ, Snipe JR, Resnick MA (1996) A double-strand break within a yeast artificial chromosome (YAC) containing human DNA can result in YAC loss, deletion or cell lethality. *Molecular and Cellular Biology* 16(8):4414–4425.
48. Rattray AJ, McGill CB, Shafer BK, Strathern JN (2001) Fidelity of Mitotic Double-Strand-Break Repair in *Saccharomyces cerevisiae*: A Role for SAE2/COM1. *Genetics* 158(1):109–122.

49. Storici F, Snipe JR, Chan GK, Gordenin DA, Resnick MA (2006) Conservative Repair of a Chromosomal Double-Strand Break by Single-Strand DNA through Two Steps of Annealing. *Molecular and Cellular Biology* 26(20):7645–7657.
50. Zwolshen JH, Bhattacharjee JK (1981) Genetic and biochemical properties of thialysine-resistant mutants of *Saccharomyces cerevisiae*. *Journal of General Microbiology* 122(2):281–287.
51. Wang Q, Wang L (2008) New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast. *Journal of the American Chemical Society* 130(19):6066–6067.
52. Wang Q, Wang L (2012) Genetic Incorporation of Unnatural Amino Acids into Proteins in Yeast. *Unnatural Amino Acids*, Methods in Molecular Biology., eds Pollegioni L, Servi S (Humana Press), pp 199–213. Available at: http://link.springer.com/protocol/10.1007/978-1-61779-331-8_12 [Accessed January 6, 2013].
53. Braglia P, Percudani R, Dieci G (2005) Sequence Context Effects on Oligo(dT) Termination Signal Recognition by *Saccharomyces cerevisiae* RNA Polymerase III. *Journal of Biological Chemistry* 280(20):19551–19562.
54. Acker J, Conesa C, Lefebvre O Yeast RNA polymerase III transcription factors and effectors. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. Available at: <http://www.sciencedirect.com/science/article/pii/S1874939912001733> [Accessed January 6, 2013].
55. Deltcheva E, et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471(7340):602–607.
56. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*. Available at: http://www.nature.com.ezp-prod1.hul.harvard.edu/nbt/journal/vaop/ncurrent/full/nbt.2508.html?WT.mc_id=TW_T_NatureBiotech [Accessed February 6, 2013].
57. Mumberg D, Müller R, Funk M (1994) Regulatable promoters of *Saccharomyces cerevisiae*: comparison of transcriptional activity and their use for heterologous expression. *Nucleic Acids Research* 22(25):5767–5768.
58. Schiestl RH, Gietz RD (1989) High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Current Genetics* 16(5-6):339–346.
59. Burland TG (2000) DNASTAR's Lasergene sequence analysis software. *Methods in Molecular Biology* 132:71–91.

60. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10(3):R25.
61. Ryan OW, et al. (2014) Selection of chromosomal DNA libraries using a multiplex CRISPR system. *eLife Sciences*:e03703.
62. Wang H, et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*. Available at: [http://www.cell.com/abstract/S0092-8674\(13\)00467-4](http://www.cell.com/abstract/S0092-8674(13)00467-4) [Accessed May 7, 2013].
63. Li J-F, et al. (2013) Multiplex and homologous recombination-mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nature Biotechnology* 31(8):688–691.
64. Bugreev DV, Rossi MJ, Mazin AV (2011) Cooperation of RAD51 and RAD54 in regression of a model replication fork. *Nucleic Acids Research* 39(6):2153–2164.
65. Mazin AV, Alexeev AA, Kowalczykowski SC (2003) A Novel Function of Rad54 Protein Stabilization of the Rad51 Nucleoprotein Filament. *Journal of Biological Chemistry* 278(16):14029–14036.
66. Caesar R, Blomberg A (2004) The Stress-induced Tfs1p Requires NatB-mediated Acetylation to Inhibit Carboxypeptidase Y and to Regulate the Protein Kinase A Pathway. *Journal of Biological Chemistry* 279(37):38532–38543.
67. Meena RC, Thakur S, Nath S, Chakrabarti A (2011) Tolerance to thermal and reductive stress in *Saccharomyces cerevisiae* is amenable to regulation by phosphorylation–dephosphorylation of ubiquitin conjugating enzyme 1 (Ubc1) S97 and S115. *Yeast* 28(11):783–793.
68. Parrini MC, Bernardi A, Parmeggiani A (1996) Determinants of Ras proteins specifying the sensitivity to yeast Ira2p and human p120-GAP. *EMBO Journal* 15(5):1107–1111.
69. Burt A, Koufopanou V (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Current Opinion in Genetics & Development* 14(6):609–615.
70. Macreadie IG, Scott RM, Zinn AR, Butow RA (1985) Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. *Cell* 41(2):395–402.
71. Windbichler N, et al. (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. *Nature* 473(7346):212–215.

72. Chan Y-S, Naujoks DA, Huen DS, Russell S (2011) Insect population control by homing endonuclease-based gene drive: an evaluation in *Drosophila melanogaster*. *Genetics* 188(1):33–44.
73. Chan Y-S, Huen DS, Glauert R, Whiteway E, Russell S (2013) Optimising Homing Endonuclease Gene Drive Performance in a Semi-Refractory Species: The *Drosophila melanogaster* Experience. *PLoS One* 8(1):e54130.
74. Takeuchi R, Choi M, Stoddard BL (2014) Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. *Proceedings of the National Academy of Sciences of the United States of America* 111(11):4061–4066.
75. Gratz SJ, et al. (2013) Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics*. Available at: <http://www.genetics.org/content/early/2013/05/20/genetics.113.152710> [Accessed September 9, 2013].
76. Friedland AE, et al. (2013) Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nature Methods* 10(8):741–743.
77. Jiang W, et al. (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Research*:gkt780.
78. Hwang WY, et al. (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* 31(3):227–229.
79. Esvelt KM, Smidler AL, Catteruccia F, Church GM (2014) Concerning RNA-guided gene drives for the alteration of wild populations. *Elife*:e03401.
80. Oye KA, et al. (2014) Regulating gene drives. *Science* 345(6197):626–628.
81. Chamberlain N, Cutts NS, Rainbow C (1952) The Formation of Pigment and Arylamine by Yeasts. *Journal of General Microbiology* 7(1-2):54–60.
82. Mao X, Schwer B, Shuman S (1996) Mutational analysis of the *Saccharomyces cerevisiae* ABD1 gene: cap methyltransferase activity is essential for cell growth. *Molecular and Cellular Biology* 16(2):475–480.
83. Tomar P, et al. (2013) Sporulation genes associated with sporulation efficiency in natural isolates of yeast. *PLoS One* 8(7):e69765.
84. Kondo S, Ueda R (2013) Highly improved gene targeting by germline-specific Cas9 expression in *Drosophila*. *Genetics* 195(3):715–721.

85. Deredec A, Godfray HCJ, Burt A (2011) Requirements for effective malaria control with homing endonuclease genes. *Proceedings of the National Academy of Sciences of the United States of America* 108(43):E874–E880.
86. Ito J, Ghosh A, Moreira LA, Wimmer EA, Jacobs-Lorena M (2002) Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature* 417(6887):452–455.
87. Li J, et al. (2013) Genome-block expression-assisted association studies discover malaria resistance genes in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America* 110(51):20675–20680.
88. Jain S, et al. (2009) A recombination execution checkpoint regulates the choice of homologous recombination pathway during DNA double-strand break repair. *Genes and Development* 23(3):291–303.
89. Costantino L, et al. (2014) Break-Induced Replication Repair of Damaged Forks Induces Genomic Duplications in Human Cells. *Science* 343(6166):88–91.
90. Looke M, Kristjuhan K, Kristjuhan A (2011) Extraction of Genomic DNA from Yeasts for PCR-Based Applications. *Biotechniques* 50(5):325–328.
91. Spitzer M, Wildenhain J, Rappsilber J, Tyers M (2014) BoxPlotR: a web tool for generation of box plots. *Nature Methods* 11(2):121–122.
92. Vyas VK, Barrasa MI, Fink GR (2015) A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families. *Science Advances* 1(3):e1500248–e1500248.
93. Gantz VM, Bier E (2015) The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations. *Science* 348(6233):442–444.

