Leveraging the awesome power of *Saccharomyces* to probe the evolution of genetic networks.

Taylor Katherine Wang

A dissertation

Submitted in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Maitreya J Dunham, Chair

Michael Ailion

M. K. Raghuraman

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Leveraging the awesome power of *Saccharomyces* to probe the evolution of genetic networks.

Taylor Katherine Wang

Chair of the Supervisory Committee:

Maitreya J Dunham

Department of Genome Sciences

Causal loci for phenotypes of interest do not exist in isolation, but rather in complex, densely connected groups termed genetic interaction networks. Given the degree of functional conservation shared with humans, *S. cerevisiae* represents an excellent model organism for understanding the connection of genetic networks to complex traits. Within the field, questions remain about the conservation of genetic networks among various, distantly related species. To address this, I studied how functional rewiring has evolved in *S. cerevisiae* and its most distantly related *Saccharomyces* relative, *S. uvarum*. I used differential gene essentiality as my phenotype of interest, which was previously observed to occur for 12% of comparable orthologs.

Differentially essential orthologs between *S. cerevisiae* and *S. uvarum* are capable of functional complementation, pointing towards changes in interaction partners as the genetic basis. My work

focused on identifying these evolutionary changes in the genetic network between species by mapping these new interaction partners. Focusing on the differentially essential orthologs that are essential in *S. cerevisiae* only, for select candidate genes I leveraged the power of transposon mutagenesis and whole genome sequencing to perform genome wide reciprocal hemizygote screen in these hybrids to map loci with differential phenotypic results depending on the identity of the mutated parental allele. Analysis of transposon insertion rates across both parental genomes in the population will make it possible to identify novel genetic interaction partners that have been gained or lost for the non-essential *S. uvarum* ortholog. Mapping the genetic network changes among these two 20 mya diverged *Saccharomyces* yeasts will provide new insight into the level of conservation across the Eukaryotic tree of life.

Acknowledgements

This defense has been a long time coming, and there are a great many wonderful people who have made it possible for me to be here today. First, I would like to thank my family for letting me explore my fascination with the natural world through acorn mudpies and caterpillar farms. I would like to especially acknowledge my father, for driving me one hour each way to begin attending college classes in high school before I could earn my license. I would like to acknowledge the first person to teach me about genetics, Ms. Beth Chagrasulis at the Lake Region High School in Naples, Maine. Ms. Shag began a life long obsession in the heart of a small town country girl that would change my life forever. I would like to acknowledge all of my wonderful community college professors who do amazing work with underprivileged students on a fraction of the budget of R1 faculty. They all encouraged my curiosity and indulged my hyper fixation on biology, chemistry, and genetics. I would like to acknowledge all of the faculty who continued their work after I transferred to UCLA. The classroom has always been a sacred, peaceful place in my life, and it is thanks to all of these teachers' collective efforts. I would especially like to acknowledge Guillaume Chanfreau, who took a chance on an 'elderly' Senior undergraduate researcher. A very special thank you to my graduate student mentor Domagoj Hodko, who truly helped spark my deep love of bench work and classical yeast genetics. I would also like to acknowledge Sri Kosuri for accepting me into his lab for my Master's thesis, under the initial guidance of Rocky Cheung who patiently taught me all I know about Gibson assembly. In addition, I want to specially acknowledge Maitreya Dunham for taking a chance on a post general exam, labless graduate student. For her unwavering belief and support in me and my ability as a scientist, I will always be grateful. Maitreya is truly a wonderful mentor and person, and being in her lab has changed my life for the better. I would like to give a heartfelt

acknowledgement to all the wonderful people I have shared the Dunham lab space with: Chris Large, Clara Amorosi, Anja Ollodart, Abigail Keller, Noah Fredstrom, Cindy Yeh, Bryce Taylor, Renee Geck, Leah Anderson, Joe Armstrong, Gabby Ferra, Barbara Dunn, Annamarie Steed, and many more. A special acknowledgement for Emily Mitchell, our wonderful lab manager that keeps us all afloat and everyone sane. I would also like to acknowledge my wonderful committee members for always offering their sage advice, and patiently listening to my slides: Michael Ailion, Aimée Dudley, M.K. Raghuraman, and Jesse Zalatan. I would distinctly also like to acknowledge that Jesse was with me even in my first committee during my time in the Fields lab, and he greatly helped smooth my lab move by being a familiar face in all of the changes. I would also like to acknowledge all of my beloved chosen family that have kept me sane in graduate school: Courtnee Clough, Sean Evans, Cindy Wei, Roger Tu, Bryan Andrews, Brooke Angell, Luana Paleologu, Daniel Reveles, Nick Woodall, Malia Marstaller, Becca Martin, Elena Romero, Sophie Moggridge, Steve Shen, Summer Sturges, Kari O'Neill and Kendra Cruickshank. All of these wonderful humans have supported me throughout all of these years and have provided great stress relief in the form of laughter and love. Finally, I would like to acknowledge my husband and life partner John Wang. Yet again, science has changed my life for the better because we met during a teaching assistant workshop at UCLA during my Master's year and his first year of graduate school. John has always encouraged me to pursue my dreams, and even followed me to Seattle at great personal effort by ending his own graduate studies there to begin again at UW. John is the most loyal, kind, and loving person I know, and it is because of his unfailing support and never-ending love that I have made it to this academic finish line.

Dedication

I dedicate this work to everyone who encouraged me to persist.

Chapter One: Introduction

1.1 Dissecting the genetic architecture underlying complex or simple phenotypes

Work over the past half a century has shown that disentangling the connection from genotype to phenotype is a convoluted problem[1,2]. Tracing genetics through history has shown a wide array of approaches: selective breeding of pea plants to discern patterns of inheritance for physical traits, generation of pools of mutants through UV irradiation and elucidation of phenotypes afterwards, to finally the isolation of an underlying gene for specific phenotypes like pink colony color in budding yeast[3]. Over the decades, the scientific community has advanced from selecting a phenotype of interest and attempting to discern the genetic basis, to purposeful and targeted generation of mutants for a gene of interest and assaying their phenotype to further elucidate function.

As approaches to understanding genotype to phenotype connections have grown in leaps and bounds, so too has the scientific community's understanding of the complexity of functions within the cell. Groups have dedicated decades of research to generating methodical knock out lines within organisms, perturbing function of genes one by one and assaying phenotype to understand the role each plays within the cell. Others have focused more recent efforts on detailed analysis of the specific gene underlying phenotypes of interest: generating libraries for each gene containing all possible amino acid replacements to study in fine detail the connection of genotype to phenotype in an approach called Deep Mutational Scanning (DMS)[4–8]. These approaches yield a wealth of information at an unprecedented level of detail, but only for one gene and often only for a specific function. However many genes, whether their final gene product be RNA or protein, have many overarching functions within a single cell and even

multiple interactions with other gene products performing their own functions[9–13]. The cell is a crowded, frantically busy place and methods of study aimed at genotype to phenotype understanding have been evolving to reflect this perception[14–17].

All of this work collectively has led to a broader comprehension that causal loci for phenotypes of interest do not exist in isolation within the cell, but rather in complex, densely connected groups termed genetic interaction networks[18–21]. While some interacting partners within a genetic network have physical interactions between their gene products, such as in a protein complex, that is not necessary for a pair of genes to be defined as connected nodes within a genetic network[22,23]. In fact, it is not even very common according to recent studies. Instead, it is commonly found that two genetic interaction partners possess overlapping or similar functions within biological pathways such as DNA damage repair and oxidation stress response[23–25]. Previous work has shown that perturbations to genetic networks in humans can cause disease, especially for genes that serve as "hubs" within genetic networks[26–28]. These are genes that are involved in many different necessary processes, or which are very important for fundamental biological processes. In addition, work interrogating the role of genetic networks in cancer phenotypes has shown that these central hubs of networks are excellent therapeutic targets for potential chemotherapy drugs and often represent the genes essential for an organism's viability[29].

In order to understand more about the connection of genetic networks to human health and disease, work has begun to generate maps of genetic networks in humans; however, there exist significant technical challenges. Most genetic network mapping approaches rely on defining

pairwise interactions via knockouts or other styles of genetic perturbation. Two genes are considered to have an interaction within a genetic network if the loss of just one or the other has a different fitness effect compared to loss of both. This deviation from expected multiplicative fitness is also often spoken about in literature under the term epistasis as well. The two terms and bodies of study are highly connected and ask many similar questions: for ease of understanding I will speak mostly within genetic network terminology.

This style of study is difficult to perform in diploid genomes such as humans because there exists a second copy of each pair of query genes such that barring haploinsufficiency or dominant negative effects, the second copy can cover the loss or dysfunction of the perturbed copy and nothing is learned about the presence or absence of an interaction between the query genes[30–33]. While some work has begun focusing on cleverly engineered haploid human cell lines, these cell lines have their own technical issues especially when attempts to extrapolate data to diploid cell lines are made.

Serious concerted efforts have also been made in key model organisms, including *Caenorhabditis elegans* and the fission yeast *Schizosaccharomyces pombe*[31,34]. Evolutionary geneticists hypothesize that there may exist a core, conserved genetic network among *Eukarya* composed of what most would consider key genes underlying the most basic and important cellular processes. In searching for a potential conserved genetic network, any comparisons between genetic networks of homologous genes between organisms are complicated by experimental issues preventing clear interpretation of genetic network comparisons. These experimental issues originate from the inability to have direct comparisons due to variation in the

specific genetic tools applied to discern each genetic network. For example, the genetic network maps generated for *C. elegans* do not rely fully on gene knockouts and must instead also leverage clever tools such as RNA interference (RNAi) or other versions of inducible knockdowns[21,35–37].

Fortunately, a large body of work over the past decade in the budding yeast *Saccharomyces cerevisiae* has generated an almost complete genetic map and provided insight into general principles of biological genetic networks[24,29,38]. Despite large differences in cell organization, metabolisms, and motility between the billion year diverged *S. cerevisiae* and *Homo sapiens*, there exists a high degree of functional conservation between the two species[39]. Fundamental processes such as DNA replication and cell cycle progression were discerned by leveraging yeast as a model organism, underlining the level of conservation[40,41]. In addition, this conservation can also be seen through the lens of essential genes required for normal cellular proliferation. When compared to 20% for non-essential genes, approximately 38% of essential genes in *S. cerevisiae* have identifiable homologs in the human genome. With this level of conservation in mind, *S. cerevisiae* represents an excellent model organism for understanding the connection of genetic networks to complex traits.

However, questions remain about how conserved these genetic networks are among various, distantly related species. Preliminary work has provided conflicting answers, but does suggest the existence of a conserved, Eukaryotic genetic network[21,21,32,37]. There have been no such comparative studies in more closely related organisms aiming to understand how genetic interactions rewire over shorter evolutionary timescales. To address this question, I have focused

the bulk of my thesis research addressing the questions of how functional rewiring of biological processes has evolved in *S. cerevisiae* and one of its most distantly related *Saccharomyces* relatives, *S. uvarum*. In particular, I have worked to map the underlying genetic architecture that gives rise to the complex phenotype of gene essentiality, which differs for hundreds of orthologs between these species.

Differential essentiality in these two *Saccharomyces* yeasts offers a way to gain insight into the evolutionary pressures on genetic networks, by gauging the level of rewiring between more closely related species than have previously been analyzed. Moreover, even though changes in genetic networks have been mapped on larger scales, sparser work has been done to understand the underlying genetic architecture of these changes at the molecular level. In order to understand genetic network evolution, gene essentiality offers an excellent, strong phenotype to study because previous work indicates essential genes are more likely to exist as important hubs of dense connections[23,26,29].

Gene essentiality is also an excellent starting point to understanding changes in genetic networks over evolutionary time because network mapping done in *S. cerevisiae* was originally done by screening for synthetic lethality[42]. A synthetic lethal screen involves taking a strain with one gene deleted, and generating secondary knockouts of genes across the genome, i.e. a one-by-all genetic screen. An array of otherwise isogenic strains is eventually generated that each has the deletion of the gene of interest, and a second deletion of a different gene[42,43]. The two genes in each double knockout strain are determined to have a genetic interaction when the phenotypic outcome of the double gene knockouts is different from their expected fitness. This expected

fitness is calculated from the observed fitness effects of both genes in separate, individual knockout strains. In the most extreme cases, a strain may be observed to be viable upon deletion of either of the two genes alone, but inviable when both are attempted to be knocked out in a single haploid strain. Thus the term synthetic lethality: by deleting the first gene a "synthetic" gene essentiality exists such that the second gene has now become essential in the mutant background.

Synthetic lethality screens are also an excellent reminder of how important genetic background is to all genotype to phenotype studies. As powerful as studies such as DMS are for deeply interrogating the sequence to function relationship of a gene of interest, they are intrinsically impacted by the genotype of the organism they are performed in. Pre-existing standing variation, for example, in the interacting genes within the genetic network of the gene of interest can drastically change the phenotypic result. The sometimes subtle effects of genetic background on phenotypic outcome for a novel mutation can also lead to incorrectly assigning phenotypic differences only to changes in the gene of interest, rather than also as an accumulation of the effect this cryptic genetic variation has on the phenotype.

In addition, as more and more DMS data sets are made publicly available to the community at large, the comparison of data sets becomes of greater interest. However, such comparisons could be confounded by strain or cell line specificity: even if two separate DMS datasets are both generated in *S. cerevisiae*, the choice of different strains, or isolates, could pose a significant problem to overcome in order to draw meaningful conclusions from comparisons[44,45]. In order to empower these sorts of studies and comparisons, the effect of the strains and their differing

genetic backgrounds on such genotype to phenotype studies needs to be further explored to be better understood. The understanding of genetic background has already grown in leaps and bounds, and its classification more recently as interesting genetic variation worth studying itself, rather than just a nuisance to be dealt with, paves the way to further deepening our understanding of its phenotypic consequences and biological significance[46,47].

1.2 The underpinnings of genetic background within the phenotype of gene essentiality between two *Saccharomyces* species.

Since its existence was first described, genetic background has generally been understood as the observation that spontaneous or induced mutations result in a differing array of phenotypic consequences among genetically distinct individuals within a species' population. In humans specifically, the effect of genetic background can be described as the effect of existing standing variation between individuals in a population on the consequence of a novel mutation. After many years of work in the fields of genetics, genomics, and evolutionary biology this phenomenon has been observed repeatedly: from the effect of a mutation in different *E. coli* strains, to the breeding of an engineered mouse line with separate inbred mouse lines yielding different phenotypic results.

The influence of genetic background on subsequent mutations can often most clearly be seen in examples pertaining to human health and disease. Cystic fibrosis as a disease model is an especially good example, because the vast majority of affected individuals are homozygous for a variant of the cystic fibrosis transmembrane conductance regulator (CFTR) protein labeled p.Phe508del[48,49]. While a group of individuals with cystic fibrosis may all have this mutation, a

large array of disease severity is observed among those individuals. Thus, there is something modulating the phenotype of cystic fibrosis severity within the genomes of these different individuals. This sort of background genetic variation is referred to as cryptic genetic variation within the field, and is a common theme, and problem, of research focused on understanding the genetic basis for human disease.

The importance of genetic background is also a pressing issue within the field of genetics as it pertains to the assignment of a 'wild type'. If there is existing genetic variation within a population such that the same novel mutation leads to differing phenotypic changes, how does one benchmark such an experiment? How does one properly average, or choose among, this sort of variation to create a heathy, "normal" individual as a control for an experiment? For example, in the past many studies focused on *S. cerevisiae* genetics and evolution leveraged the laboratory strain named S288C. While S288C is often treated as a ubiquitous control, the field now understands there exists a significant level of sequence and phenotypic variation present in this strain compared to natural isolates and even other laboratory strains[46,50]. Examples of this sequence and phenotypic variation including loss of function (LOF) mutations in important genes involved in the speed of sporulation and meiosis.

The concept of a single 'wild type' individual and its consequences is also of particular interest within the field of human health and disease. In an attempt to address the shortcomings of choosing a single person or few persons' genomes as point of comparison for new data, a new research mission was formed (the All of Us initiative) by the NIH focused on sequencing one million individuals belonging to many different backgrounds and areas of the world[51]. By

considering the effect of genetic background on phenotypic outcomes for a mutation, the community enriches their understanding of disease severity in different populations while also beginning to dismantle the idea of a single, fixed 'wild type' in human genetic studies.

While the field continues to grow and expand as more work interrogates the molecular underpinnings of genetic background effects, certain patterns have been observed. Thus far, the general understanding is that the phenotypic variation seen in background effects upon a newly acquired mutation can be understood through the lens of these genetic background effects representing genetic interactions between the mutation and segregating loci. This understanding of genetic background effectors as genetic interactors with the mutation of interest has been especially well studied within the *Saccharomyces* community[14,52,53]. The relative ease of genetic mapping studies such as QTL has made it possible to observe just how much variation in these interacting loci under the genetic background umbrella can affect or interact with a genotype of interest. For example, mapped segregating loci that interact with the novel mutation and affect phenotypic expression can be found to themselves interact with each other in a small, local genetic network. The variation in genetic backgrounds is thought to arise from mutations at these interacting or epistatic loci. These mutations can range from affecting core functional processes like signaling cascades and regulatory pathways, to being part of vastly different cellular processes. Even with this knowledge, these general patterns of observed molecular mechanisms of genetic background effects are often rather specific on a case by case basis, and dependent on the function and phenotype of the novel mutation whose phenotypic expression is altered by the genetic background.

One such example of this comes in the form of studying essentialomes—the most complete list of essential genes known and profiled in the organism of interest—among individual isolates or strains. The term essential gene seems self explanatory, but variation in lists of essential genes has been observed in different genetic backgrounds of *S. cerevisiae* for quite a while. Exploratory work into the molecular mechanisms underpinning these differential essentiality phenotypes between the common laboratory strain S288C and another isolate Σ1278b has shown that the answer is not overly simple[54]. There does not seem to be a single overarching driver or controller for differentially essential genes between strains, and that instead it was found that changes in local gene-gene interactions at differing levels of complexity could provide explanations. The question remains: are these causal local gene-gene interaction changes the standard for what explains differential essentiality, or are these cases specific and niche to the variation between each pairwise comparison between S288C and Σ1278b, or even any other isolate? In addition, the task of profiling an essentialome for a specific species or strain is not quite so simple an undertaking, making these kinds of comparisons challenging, but ultimately informative.

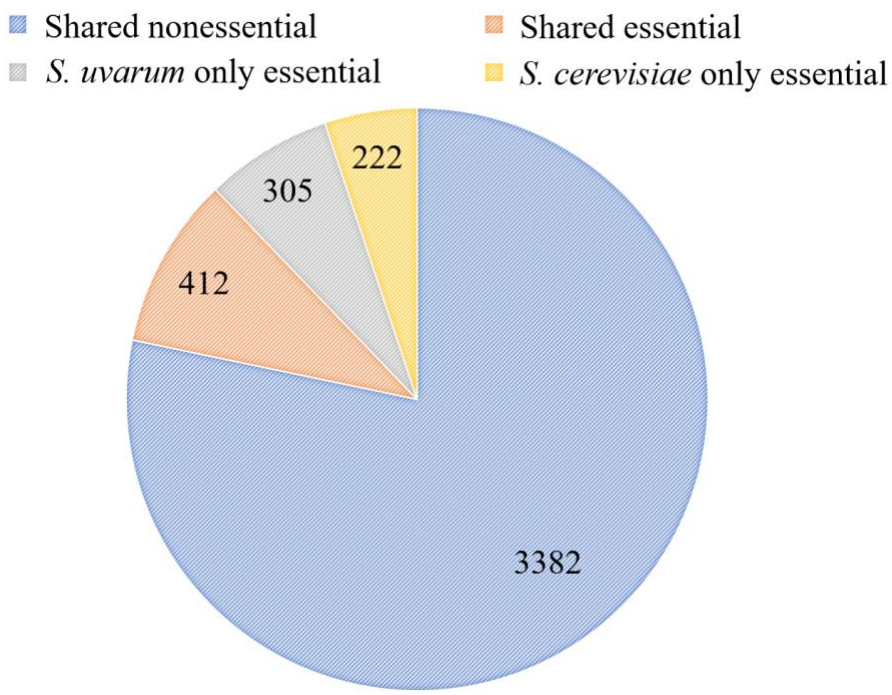
The definition of genetic background so far has been discussed as the understanding that variation exists in separate individuals within the same species. However, taken to an extremum the concepts of genetic backgrounds can be applied to understanding the differences between two closely related species with conserved and shared biological processes. Specifically, my thesis work was focused on applying the framework for how genetic background affects gene essentiality between different *S. cerevisiae* strains, to understanding changes in gene essentiality between *S. cerevisiae* and one of its most distantly related *Saccharomyces* relatives, *S. uvarum*[55–

[58]. By leveraging a similar approach as was used to interrogate the genetic basis for differential gene essentiality between *S. cerevisiae* isolates, I aimed to ask a similar question but across a much larger evolutionary timescale. A potential exists that only small, local changes were observed in these studies with *S. cerevisiae* isolates because of the smaller evolutionary space explored by the genetic background effect studies so far, and limitations of statistical power within past genetic methods. Therefore, expanding differential gene essentiality studies beyond genetic background effects within a species and instead exploring between species could empower a study to discern any larger drivers of the effects of genetic backgrounds on the phenotype of gene essentiality across a larger evolutionary time scale.

This work was not even possible until quite recently. While decades of work have made it possible to functionally characterize the *S. cerevisiae* genome and its essential genes, the lack of equivalent tools and resources in non-model organisms was previously a barrier to understanding functional genome evolution across the *Saccharomyces* genus[59]. As part of overcoming this challenge, the Dunham lab previously applied an *in vitro* transposon mutagenesis assay to functionally characterize the genome of *S. uvarum*[55]. Leveraging this process, a list of predicted essential genes was generated for both *S. cerevisiae*—which was comparable to preexisting literature defining the essentialomes—and *S. uvarum*.

Comparing these dispensability profiles across the 4,321 orthologous gene pairs for *S. cerevisiae* and *S. uvarum* revealed that 12% of orthologs had differential dispensability [Figure 1.2.1.]. The essentialome changes point towards interesting functional genomic differences between the two

Figure 1.2.1. Pie chart showing the distribution of predicted gene essentiality between comparable orthologs of *S. uvarum* and *S. cerevisiae.*



species arising from the divergence of gene function and/or changing genetic network interactions. There were two potential explanations for the observed differential gene essentiality: either function of the orthologs had diverged significantly enough since last common ancestor that the orthologs were performing different functions, or the gene-gene interactions in the non-essential species had changed relative to the ancestral state. This hypothesis of genetic network changes underlying species specific genes arises because another way of considering differential gene essentiality is as suppressors in the non essential species acting through an altered or novel interaction between the suppressor and the differentially essential gene in the tolerant species.

To discern between these two hypotheses, a previous member of the lab performed cross-species complementation studies. She observed that when a *S. cerevisiae* strain heterozygous for an essential gene knockout was transformed with a 'wild type' copy of the *S. uvarum* ortholog, the resulting haploid containing the gene deletion and the *S. uvarum* ortholog recovered after

sporulation and tetrad dissection was viable [Figure 1.2.2.]. This cross-species functional complementation pointed to the other hypothesis as an explanation of the differential gene essentiality: changes in local genetic networks, made up of interacting loci with the gene of interest, between *S. cerevisiae* and *S. uvarum* that have accumulated since their last common ancestor. Based on the complementation results, we hypothesized that the underlying genetic bases of these differential dispensability profiles between *S. cerevisiae* and *S. uvarum* were due to changes in the involved genetic networks and represent a sort of extremum of genetic background effects.

The work discussed here aimed to address the hypothesized, underlying changes in genetic architecture by performing genetic mapping to understand the molecular mechanisms underpinning these changes in gene dispensability profiles between the two related yeast species previously discussed: *S. cerevisiae* and *S. uvarum*. Several features make *S. uvarum* an excellent organism in which to perform these comparative studies: it is genetically tractable using methods developed in *S. cerevisiae*, *Saccharomyces* yeasts have the useful ability to hybridize making cross-species genetic mapping possible, and *S. uvarum* is one of the two most distantly related *Saccharomyces* yeast to *S. cerevisiae*[56]. These features of *S. uvarum* make the proposed genetic mapping experimentally feasible in this non-model organism, while also preserving the ability to discern and map changes in genetic networks between two related species that diverged ~20 million years ago and interrogate the effect of genetic background on gene essentiality across a larger time scale than previously examined[60]. In addition, mapping these rescue loci will inform the field's understanding about how genetic networks evolve, and the role genetic background

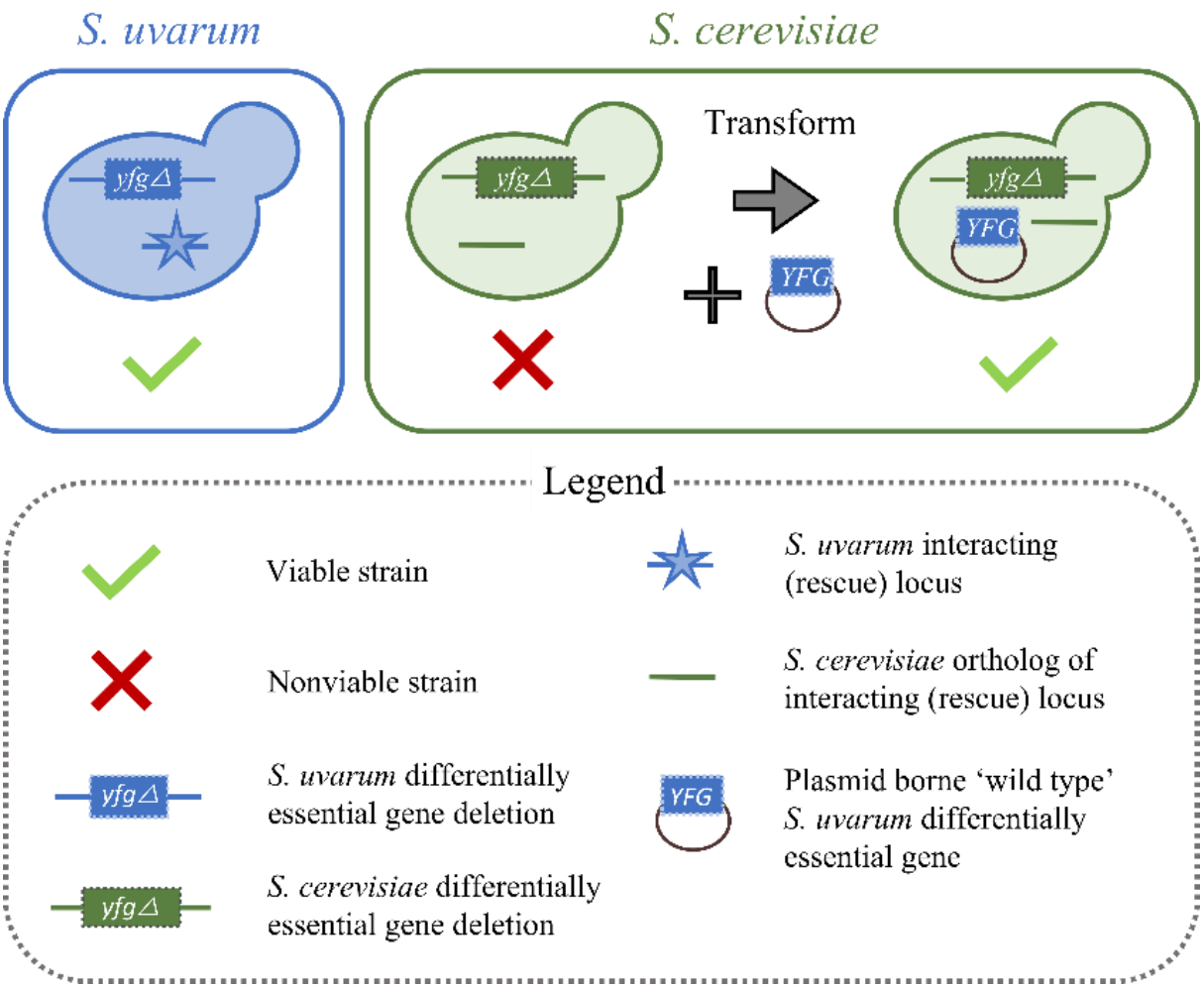


Figure 1.2.2. Diagram of cross-species complementation assay showing the rescue of an *S. cerevisiae* species specific essential gene loss by the 'wild type' *S. uvarum* ortholog.

plays across differing timescales.

Work towards understanding the genetic determinants of differential gene essentiality between *S. cerevisiae* and *S. uvarum* will be critical not only for mapping the changes in genetic networks between the species-specific essential genes, but also for generating a greater understanding of how this higher level of biological organization evolves. Understanding these fundamental principles will be necessary in order to chart a path from the genetic interactions gleaned in *S.*

cerevisiae and how they relate to genetic networks present in humans[19,33]. This long term goal is especially important for networks underlying complex human disease, since targeting hubs of genetic networks is a viable approach for finding cancer drug targets[26].

1.3 Barriers to traditional genetic mapping approaches for *S. cerevisiae* x *S. uvarum* interspecies hybrids

Previous work studying genetic background effects in yeast described the number of interacting loci—the genetic architecture—underlying these effects to vary on a gene by gene basis. Given this variation previously observed, we expected that the genetic architecture underlying species-specific gene essentiality would be of varying complexity, representing a potential pitfall. Even with a traditional en masse QTL based approach named Bulk Segregant Analysis (BSA) that has greater statistical power than original methods, mapping multiple loci of small effects will be challenging[61–63]. To overcome the challenges of mapping many loci of small effect, *S. cerevisiae* strains from the Synthetic Genetic Array (SGA) containing strain collection can be leveraged. With a combination of mechanical and chemical methods to destroy parental cells, a significant number of viable progeny can be recovered from sporulated diploid SGA strains. Mass recovery of viable haploid offspring from a mating make it possible to perform QTL mapping for multiple strains in parallel. By leveraging SGA methods, the number of progeny required for BSA can be recovered, and this approach can be used to examine several different species specific essential gene candidates. After the subsequent mapping of the sequence basis for differential gene essentiality, I expect to observe varying levels of complexity for the underlying genetic architecture. By selecting several candidates, I increase the chances of being able to map the causative loci for at least some of these differentially essential genes even if the effect size is

small for many contributing loci.

In the case of applying QTL mapping to understanding differential gene essentiality between *S. cerevisiae* and *S. uvarum*, the parent strain subjected to screening would be a hybrid. For a candidate gene that is an essential gene specific for *S. cerevisiae*, the haploid *S. cerevisiae* partner would be 'wild type' for the differentially essential gene. The *S. uvarum* partner, on the other hand, would carry a deletion allele of its ortholog marked by a dominant drug resistance cassette. This hybrid would then be sporulated, and progeny collected for sequencing. Rather than generating multiple bins across a spectrum of phenotypes, for studying differential gene

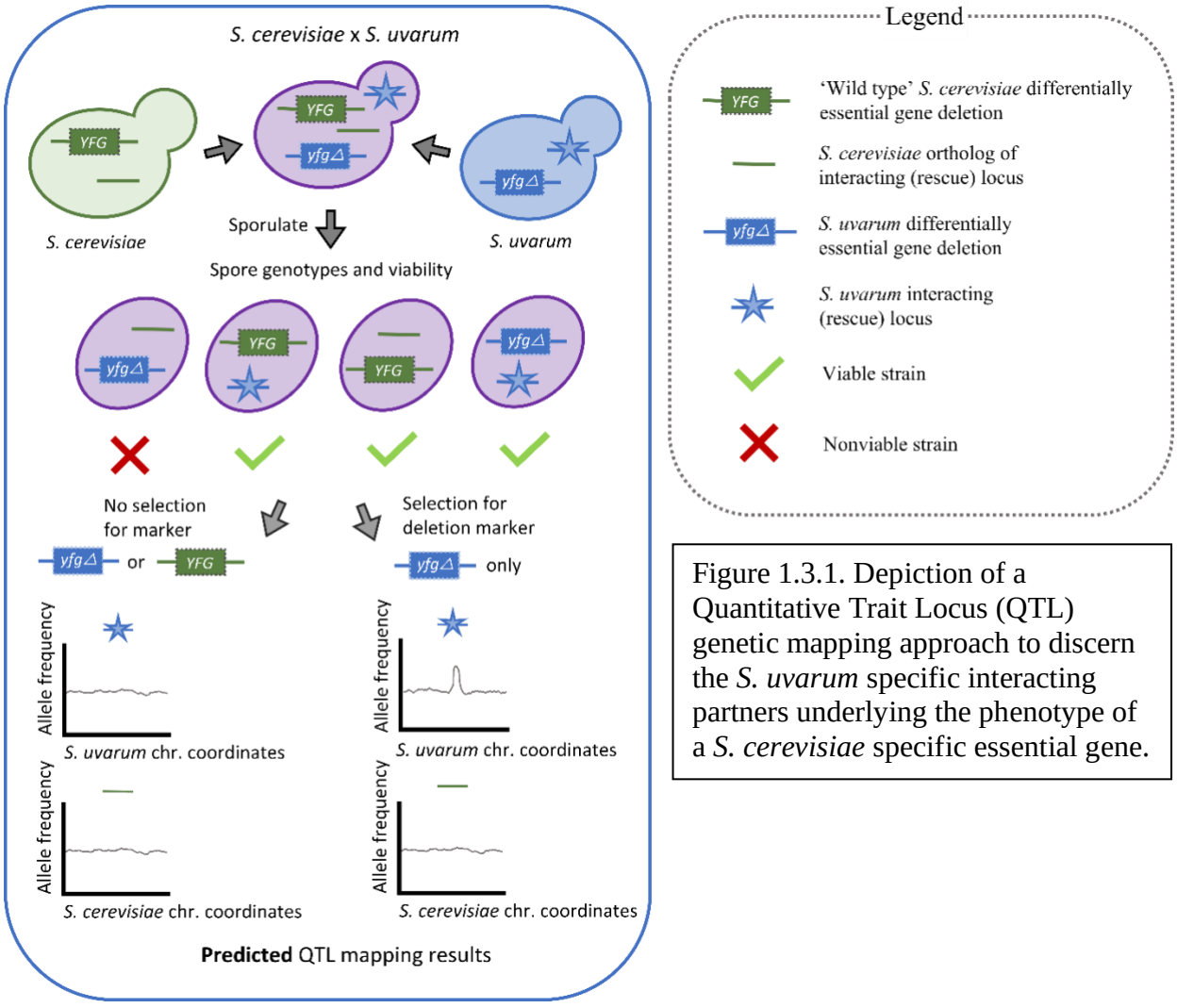


Figure 1.3.1. Depiction of a Quantitative Trait Locus (QTL) genetic mapping approach to discern the *S. uvarum* specific interacting partners underlying the phenotype of a *S. cerevisiae* specific essential gene.

essentiality the progeny of interest would have inherited the differentially essential gene deletion allele while remaining viable. The hypothesis would be that only progeny that had also inherited the interacting factor allele in the *S. uvarum* genome that suppresses the gene essentiality would be able to survive. This uniparental inheritance could be detected within the genomes of progeny of interest, when compared to a 'wild type' control, by sequencing and mapping to the parental genomes [Figure 1.3.1.]. Unfortunately, a traditional QTL strategy is a non starter as a genetic mapping approach to understand the loci underlying differential gene essentiality between *S. cerevisiae* and *S. uvarum.* The inviability of QTL mapping as an approach is because *S. cerevisiae* and *S. uvarum* are able to mate and form hybrids, but these hybrids are sterile and produce very few viable progeny[57,64,65].

Most *Saccharomyces* species are capable of mating and forming inter-species hybrids. The occurrence of hybridization has been observed both within the laboratory and also within isolates from natural populations. Hybridization often gives the resulting F1 an evolutionary advantage through the level of heterozygosity and genetic diversity generated that is then subject to natural selection[66–68]. The caveat of these hybridization events however is that F1 hybrids are largely sterile. The formation of hybrids occurs within all Kingdoms of life and sterility of the F1 progeny is common, and even in cases where F1 are fertile often there is a subsequent hybrid breakdown among later generations[69–71].

The genetic basis for hybridization barriers between two species varies between organisms—from Dobzhansky -Muller incompatibilities to ploidy issues—as well as the timing of the barrier. For example the lack of overlapping habitat, or different mating behaviors represent a pre-zygotic species barrier. An example of a post-zygotic barrier is the inability of a newly fertilized

offspring to survive to birth, or if born the hybrid is sterile and unable to produce offspring[72–75]. For the case of *Saccharomyces* hybrid infertility, genomic sequence divergence represents a post-zygotic barrier. This conclusion was the result of the work from several different groups across the last few decades[76–78]. This body of work culminated in the understanding that sequence divergence between the parental genomes in the hybrid triggers the recruitment and action of the Mismatch Repair (MMR) pathway.

MMR plays several important roles in the cell including the maintenance of genomic integrity during normal cellular growth and mitotic division, but it also plays an important role in meiosis. The 20% sequence divergence between the *S. cerevisiae* and *S. uvarum* genomes in a hybrid affects how homeologous recombination is initiated by the cell during meiosis, and results in the recruitment of MMR. After chiasma initially form, recruited MMR machinery interprets the sequence divergence between, for example, chr. I of *S. cerevisiae* and chr. I of *S. uvarum* as non-homologous chromosomes. MMR works to halt crossing over and resolve any in progress so that the "paired" parental chromosomes are unable to line up properly on the metaphase plate during meiosis I due to lack of mechanical tension created by crossing over events between them; resulting in rampant aneuploidies and inviable offspring.

The inability of *Saccharomyces* hybrids to undergo meiosis and produce viable spores provides a substantial barrier for genetic mapping approaches to elucidate the identity of rescue loci underlying differentially essential genes between *S. cerevisiae* and *S. uvarum*. Hybrid sterility especially confounds the use of traditional mapping methods such as QTL which rely on the sorting, phenotypic profiling, sequencing, and analysis of progeny from a *S. cerevisiae* x *S. uvarum* hybrid. Despite this challenge, previous work in *S. cerevisiae* x *S. paradoxus* hybrids has

shown that this sterility caused by the action of MMR can be alleviated to various degrees by mutating components of the MMR pathway.

Work from the Delneri group proved that when a factor for MMR is deleted for both *S. cerevisiae* and *S. paradoxus*, hybrids formed from these *msh2*Δ strains were able to produce more surviving progeny. Recently, the Grieg group took this approach even further by engineering meiotic null alleles at both the *MSH2* and *SGS1*—a DNA helicase involved in cross over formation—loci[79]. To avoid genome fidelity issues present in MMR deletion strains, the group instead utilized previous knowledge in the community to choose a native promoter for an endogenous gene that is specifically shut off during meiosis. The specific gene whose promoter was of interest to the Grieg group was the B-type cyclin *CLB2*, whose expression is shut off during meiosis because its protein product promotes the entry of the cell into mitosis[80,81]. By replacing the native promoters for both genes with the promoter for the gene *CLB2*, both *S. cerevisiae* and *S. paradoxus* strains were generated that specifically shut off *MSH2* and *SGS1* expression exclusively in meiosis. Results from this experiment showed that up to 40% of progeny were viable, and full tetrads were obtainable to even be able to observe a marked increase in recombination rate over wild type hybrids. This approach is a very promising one to overcome the barrier of hybrid infertility for mapping differential gene essentiality between *S. cerevisiae* and *S. uvarum*. A noted potential pitfall for this approach is that *S. cerevisiae* and *S. paradoxus* have syntenic genomes with only 10% sequence divergence. Compared to that, the 20% sequence divergence between *S. cerevisiae* and *S. uvarum* as well as several chromosomal translocations could heavily affect spore viability from *CLB2* promoter engineered *S. cerevisiae* x *S. uvarum* crosses[56].

1.4.1 Transposon mutagenesis approach to genetic mapping via reciprocal hemizygosity screen.

While QTL mapping approaches have been broadly used in many different model organisms, there exist several other quite powerful genetic methods to perform mapping. A particular method of interest that has been leveraged previously with varying success in hybridss is transposon mediated mutagenesis. Transposable elements (TEs), or transposons, represent a valuable genetic tool with many applications that have been developed since the existence of these selfish genomic elements was first described by Barbara McClintock in the middle 20th century[82]. Since their original description, extensive studies have been done upon their biology, molecular mechanisms of their jumping, and evolutionary consequences. Sometimes viewed through the lens of a selfish genetic element, or genomic parasite, transposons have been observed throughout the Kingdoms of life. Between species, transposon sequences represent different percentages of the whole genomic sequences; ranging from roughly 50% of the human genome to about 90% of the maize genome—the original species studied by McClintock and Creighton.

Originally discovered in connection to McClintock and her graduate student Creighton's work on tracking and studying genetic recombination, transposable elements represent mobile genetic elements, termed "jumping genes" capable of moving between genomic locations during cell divisions. The molecular process of this jumping within the genome can often be described as a copy and paste, or cut and paste process. This depends on whether the transposon can be classified as Class 1 Transposable Element (TE) or Class II TEs, also referred to in the literature as retrotransposons or DNA transposons, respectively[82]. The transposon first discovered by

McClintock and Creighton was named Ac/Ds, which represented a class II TE. Other notable class II TEs include the *Tc-1* / *mariner* classes of transposons, *piggybAC,* and *Sleeping Beauty*. These elements were first discovered in *Drosophila mauritiana*, the cabbage looper moth *Trichoplusia ni*, and artificially revived from a dead version in the *Salmonid* fish respectively[83–86]. DNA transposons are also referred to as complete transposons, because they encode a transposase enzyme capable of performing the cutting and pasting function responsible for their jumping within a genome. This makes the transposon independent from the main cellular machinery and does not co opt any of the endogenous proteins of the organism's genome. While all of the class II transposons listed above perform similar functions, large differences exist in the ability of the transposon to cut and paste cleanly. Within the wider family of transposons, the mechanism of action for some leave genomic signatures behind from their original locations, also termed genomic scarring, while others do not. This is of particular importance to geneticists who wish to utilize transposon as a tool, because it can affect the ability to engineer a strain or perform a mutagenesis experiment.

This self-reliant mechanism is crucial for transposition as a tool, but is different from the other class of TEs, retrotransposons, which can be further subdivided into autonomous or nonautonomous class I TEs depending on if the transposon encodes any of the proteins necessary for retrotransposition. Most autonomous retrotransposons can be further subdivided into LTR (Long Terminal Repeats) vs non-LTR transposons[87,88]. LTR transposon make further genomic changes upon pasting wherein target sequence duplication occurs, whereas non-LTR transposons create smaller, roughly 7 to 20 base pair duplications upon insertion at the target site. Both types of autonomous retrotransposon encode their own reverse transcriptases and nucleases for

transposon replication and insertion (the copy and paste mechanism). Well known examples for autonomous retrotransposons include the human endogenous retroviruses (HERV) and various Ty elements found in the genome of *S. cerevisiae*[89–91]. The discovery of transposons jump started a novel field in genetics, focused on understanding the movement and evolutionary impact of transposons and their jumping events in addition to studying their bizarre, sometimes non-Mendelian inheritance due precisely to this movement within the genome. The challenge of physical mapping for transposons within the genome was also of great interest, and part of what led to the original hypothesis of their existence by Thomas Hunt Morgan long before the tools to discern them were available.

In recent years the molecular biology, genetic, and synthetic biology fields have cleverly adapted these natural, biological elements into tools for genome engineering and genetic studies. Leveraging transposons, almost exclusively class II TEs, groups were able to co-opt cutting and pasting to perform untargeted mutagenesis beginning with simple microbes like *E. coli*. By design, the proteins encoded for by a class II transposon like *piggyBac* are instead expressed from an extrachromosomal plasmid vector, which also contains a specific sequence of interest that replaces the original transposon sequence between the sequence arms used for transposition. This sequence can vary greatly, depending on the experimental design a transposon is being used for, but is often a drug marker of some sort to tag cells that have undergone transposition where the transposon has jumped from the plasmid to the genome.

This transposon mutagenesis approach of cutting and pasting from an exogenous plasmid to genomic DNA was of particular interest to my thesis work. Transposon mutagenesis represents a

powerful genetic tool that offers an alternative mapping approach for the molecular basis of differential gene essentiality between *S. cerevisiae* and *S. uvarum*. By avoiding the need for sexual reproduction, transposon mutagenesis circumvents hybrid sterility because it can be done within the cellular context of a *S. cerevisiae* x *S. uvarum* hybrid. Specifically, in a *S. cerevisiae* x *S. uvarum* hybrid that is homozygous LOF for a species specific essential gene, transposon mutagenesis can be leveraged to perform a reciprocal hemizygosity screen for rescue factors. A reciprocal hemizygosity screen, in the most simple terms, interrogates whether there is a different phenotypic result between disrupting one parental allele or another within a hybrid. In the case of differential gene essentiality, a transposon mutagenesis based reciprocal hemizygosity screen makes it possible to map the underlying interacting loci. For a *S. cerevisiae* species specific gene, parental biases in recovered transposon insertion rates will reveal the identity of *S. uvarum* rescue loci alleles. Any loci that experience very few insertion events only within a *S. uvarum* parental allele are candidates for the loci underlying differential gene essentiality. Thus the usefulness of a reciprocal hemizygosity screen for the problem of differential gene essentiality is revealed: the parental loci in the hybrid that have different phenotypic results (viable vs inviable) upon transposon insertional mutagenesis will be candidates for the rescue loci that are being mapped.

Chapter Two: Strain engineering approach to overcoming hybrid infertility barriers to QTL mapping.

2.1 Introduction

2.1.1 History of *Saccharomyces* hybrids

A powerful feature of *Saccharomyces cerevisiae* as a model organism comes not just from the species itself, but also from its relationship to the other closely related species within the *Saccharomyces* clade. Although sequence divergence between *S. cerevisiae* and the other clade members varies by up to ~20%, *S. cerevisiae* is capable of cross-species mating with other clade members, resulting in hybrids. These hybrids exist both naturally and in domesticated settings where the pressure of natural selection is imposed by humans. There is even some bioinformatic research leading speculation that the entire *Saccharomyces* clade began with a common ancestor resulting from hybridization, followed by whole genome duplication[92–95]. The formation of a hybrid between two different *Saccharomyces* species can be an evolutionary advantage: generating genetic diversity through combining the two parent genomes into one cell. Hybrids often show genomic changes occurring at a more accelerated rate than intraspecies mating due to the inherent instability of the hybrid genome.

These are just a few of the genomic changes that can occur after hybridization in *Saccharomyces* species. There are other important consequences to hybridization as well, however. While *Saccharomyces* yeasts can form hybrids, these hybrids are sexually reproductive dead ends because they are all infertile. Hybrid infertility is not an uncommon phenotype and can be found throughout the tree of life including mammals, plants, and microorganisms such as budding yeasts. This infertility is viewed as one of many mechanisms that play a role in speciation. The

mechanisms of hybrid infertility vary based on the taxonomy of the hybrid's parents: in mammals Dobzhansky-Muller genomic incompatibilities commonly prevent most F1 hybrids from producing F2 offspring[69,96–98]. In contrast, the infertility of plant hybrids is usually caused by abnormal ploidy in the hybrid compared to the parents[99,100]. These observations were the result of decades of early study however, and the last few decades has shown that there are not firm rules and these mechanisms of speciation and hybrid infertility occur in different combinations.

2.1.2 Sterility of hybrids caused by anti-recombination activity of MMR

After the formation of yeast hybrids was discovered, they were further observed semi frequently within domesticated and wild ecological settings. Once it was determined by researchers that *Saccharomyces* hybrids were indeed infertile, the causality of this infertility became a focus of many research groups. Although there is some evidence that Dobzhansky-Muller genomic incompatibilities between parental genomes play some part in the infertility of certain hybrids, it has been hypothesized that most hybrid infertility is due to the action of the mismatch repair (MMR) pathway in these hybrids[101]. MMR machinery proteins are highly conserved across the tree of life and are critical to the cell for the role they play in the maintenance of genomic integrity[102–104]. Specifically, MMR functions to detect base - base mismatches and insertion/deletion mis-pairs generated after DNA replication and recombination.

If a cell suffers double stranded DNA breaks (DSBs) during mitosis, the DNA damage is repaired via homologous recombination (HR) with the help of the MMR pathway. HR functions to repair DSBs in a yeast cell by using the undamaged homologous chromosome as a template

via complementary base pairing to repair the DNA damage. The proteins in the MMR pathway work together to ensure homogeneity of the template and repair strand by preventing homologous recombination. If the sequences are non-homologous, such as different chromosomes or even different segments of the same chromosome, MMR prevents recombination because it would then be a case of non-homologous recombination. The action of MMR on sequence differences between a template and damaged chromosome during mitotic growth originates in the cell's need to prevent sequence from an entirely different chromosome being used to repair a DSB (i.e. chr. I instead of damaged chr. IV), rather than the sequence within the homologous chromosome[105,106]. Improper repair of a damaged chromosome can be catastrophic for the cell by the potential loss of critical genomic sequences, and thus MMR performs a necessary and critical function for maintaining genomic integrity during vegetative growth.

MMR also plays an important role in meiosis during sexual reproduction: ensuring efficacious recombination between homologous sequences between prophase I and metaphase I. The quality control provided by the pathway performs a critical function in differentiating proper cross overs between parental homologous chromosomes. When there is a certain level of sequence divergence between the recombining parental genomes, MMR views the parental chromosomes as homeologous instead of homologous. The amount of sequence divergence sufficient to trigger MMR recruitment is quite low: 1% divergence alone results in a statistically significant decrease in fertility for the parental hybrid. The pathway acts to abort the formation of any recombination intermediates and aborts ongoing intermediates by resolving the intermediates such that no crossover event will occur between the parental genomes [Figure 2.1.1][107].

Without recombination events among the parental genomes, there is not enough mechanical tension created by the physical connections of the chiasma between the homeologous chromosomes to properly line up on the metaphase plate before anaphase begins and chromosomes are distributed to daughter cells[105,108]. This is the specific mechanism underlying hybrid infertility in *Saccharomyces*: dysfunctional metaphase leading into anaphase causes massive aneuploidies in offspring thus rendering almost all of them inviable. In fact, the level of sequence divergence triggering MMR to block homeologous recombination is so little that even intraspecies diploids of various *S. cerevisiae* strains display decreased spore viability. A sequence divergence of 2% is enough to observe a reduction of spore viability from 95% to approximately 80%[67].

The lack of sexual reproduction within *Saccharomyces* hybrids is a major source of frustration for the field because it creates a considerable blockade to many of the classic genetic mapping strategies used within species. Without the genetic diversity among offspring that is generated during meiosis, it becomes immensely difficult to study genotype to phenotype connections in hybrids. Approaches like QTL mapping can be extremely powerful and would be the first choice for mapping the genetic basis of differential essentiality between *S. cerevisiae* and *S. uvarum* orthologs. However, as detailed above QTL mapping requires the production of offspring and so is impossible to use in infertile hybrids.

2.1.3 Loss of function and conditional MMR alleles to improve hybrid spore viability

A potential solution to the hybrid infertility issue was raised when mutations in the prokaryotic

MMR factors mutL, mutS, and mutH were found to greatly increase the recombination rate between *Escherichia coli* (*E. coli*) and *Streptococcus pneumonia*—up to 1000X more over wild type MMR genes[109–112]. At the time, the homologs of MutL and MutS had been identified in *S. cerevisiae* as *PMS1* and *MSH2* respectively[113]. The functional conservation with their bacterial counterparts was suggested by several observations: both *PMS1* and *MSH2* mutants in yeast had an increase in the segregation frequency of genetic markers indicative of unrepaired heteroduplex DNA, similarly to the prokaryotic homologs[114–116]. In addition, both mutants exhibit a mutator phenotype further suggesting these strains also have poorer fidelity of DNA mismatch repair in mitotic growth as well.

All of these results together prompted N. Hunter et. al to investigate the effect of MMR mutations and deletions of pathway members on recombination rate and hybrid fertility in *Saccharomyces* yeasts[108,113]. The group did so by mating laboratory *S. cerevisiae* and *S.*

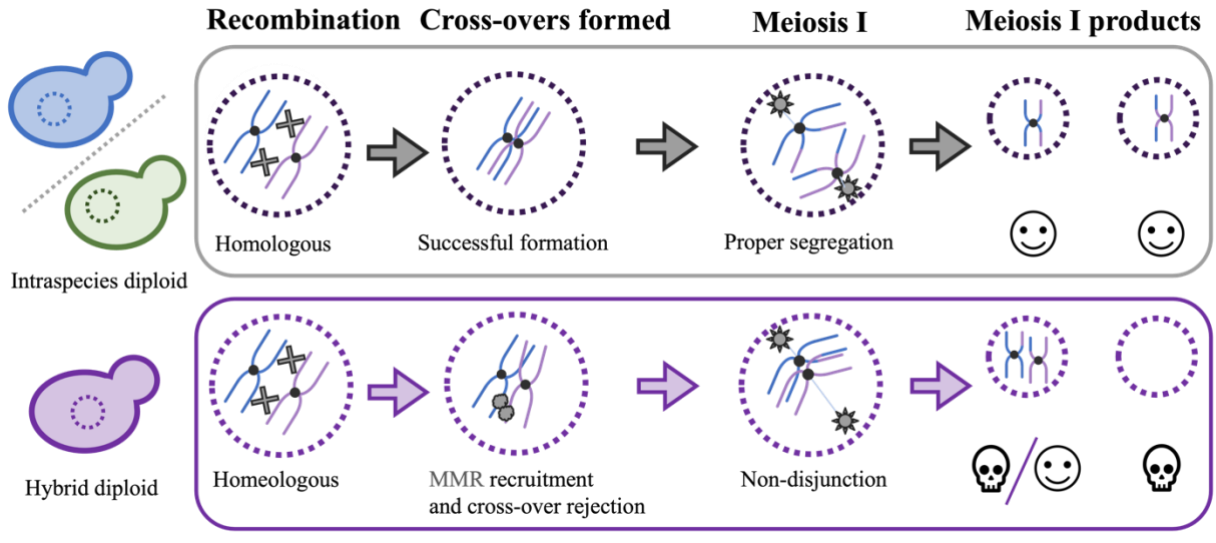


Figure 2.1.1. Visual representation of the anti-recombination activity of MMR factors resulting in poor offspring viability in inter-species *Saccharomyces* hybrids.

paradoxus strains and analyzing spore viability after sporulation. At the time, *S. paradoxus* was identified as *S. cerevisiae*'s most closely related sister species, with the genomes estimated to have ~10% divergence[115,117,118]. Of further note, karyotyping at the time had confirmed that there also existed no major structural rearrangements between the two species' genomes, suggesting extensive synteny. These hybrid strains were then sporulated to induce meiosis via nutrient starvation and their tetrad ascospores manually dissected. The rate of surviving progeny for each strain was compared to the spore viability of the 'wild type' hybrid, which was ~1%. The group found that by disrupting *MSH2* and *PMS1* individually, spore viability for the *S. cerevisiae* x *S. paradoxus* hybrid was increased to 10.2% and 7.2% respectively. In addition to this, and further supporting the important role of anti-recombination in hybrid infertility, surviving spores were found to have at least a two-fold increase in recombination events compared to wild type. These results laid the basis for continuing research that would attempt to further increase hybrid fertility.

A recent method that made a great stride in tackling this problem came from Duncan Grieg's group at University of College London[79,119]. Building upon the previous work in MMR factors and anti-recombination genes more generally, Grieg's group decided to target *MSH2,* the MutS homolog from before, and *SGS1,* a DNA helicase previously found to also be involved in preventing homeologous recombination. Their approach, however similar to previously, tried something quite inventive. As important as *msh2* mutants have been for studying the role of anti-recombination in hybrid inviability, it is critical to keep in mind that any MMR mutants exhibit a mutator phenotype. This means that in vegetative growth these strains have an increased mutation accumulation rate due to the dysfunction of MMR from null *msh2* alleles. Previous

work focused on reducing the mutation accumulation by limiting mitotic growth to only what was strictly necessary to gain enough cells to induce sporulation.

While limiting periods of mitotic growth minimizes deleterious effects from mutations in MMR machinery, in an ideal world a trade off could be avoided between hybrid spore viability and mutation accumulation. A potential way to avoid such a trade off would be to turn anti-recombination genes off specifically only during meiosis so that their critical DNA repair functions during vegetative growth could be maintained. This is exactly what the Grieg group did: they replaced the native promoters of both *MSH2* and *SGS1* with the promoter sequence for the B-type cyclin *CLB2*. *CLB2* functions generally in the cell cycle by promoting the transition from G2 to M phase, and then its protein product is specifically and efficiently marked for degradation and its expression silenced[120]. By placing both *SGS1* and *MSH2* under the *CLB2* promoter, their expression will specifically shut off only during meiosis.

The group hypothesized that without the presence of *MSH2* and *SGS1*, homeologous recombination in the hybrids would proceed and result in proper chiasmata formation and crossing over leading to viable progeny from the lack of chromosome nondisjunction. This is indeed what G. Bozdag et al. found: by creating meiotic null alleles of *MSH2* and *SGS1* under the control of the *CLB2* promoter, *S. cerevisiae* x *S. paradoxus* hybrids spore viability increased up to 70-fold. This marked increase is almost to the fertility of diverged non-hybrid crosses, and euploid offspring recovered showed multiple recombination events between each set of homeologous parental chromosomes. The drastically increased fertility and the ability to recover euploid offspring with successful genomic recombination makes it possible to perform genetic

mapping in sterile hybrids.

This approach therefore became the first method I leveraged in an attempt to perform genetic mapping in my own *S. cerevisiae* x *S. uvarum* sterile hybrids. By placing the *SGS1* and *MSH2* genes under the *CLB2* promoter for both parental species in these hybrids, I hoped to enable QTL mapping in order to uncover the genetic determinants of differential gene essentiality observed between the two species. One caveat that bears mentioning however, is that like the other groups working on anti-recombination G. Bozdag et al. performed their experiments in *S. cerevisiae* x *S. paradoxus* hybrids. As mentioned earlier, these species show one of the lowest divergences (~10%) between inter-species hybrid *Saccharomyces* parents, and the genomes are without large scale chromosomal rearrangements. This is unfortunately not true for *S. cerevisiae* x *S. uvarum* hybrids, who are on average twice as diverged at the sequence level (~20%) and have several chromosomal inversions. This difference in synteny between the hybrid parents was a cause for concern, but not reason enough to forego the method altogether.

2.2 Results

2.2.1 A homozygous *msh2Δ S. cerevisiae* x *S. uvarum* hybrid displays a small increase in spore viability.

I was interested in applying the strategy of disrupting anti-recombination machinery to my own work with *S. cerevisiae* x *S. uvarum* hybrids. First, I had to confirm these meiotic null and full deletion alleles would work in my system. This is because all the previous work on disrupting anti recombination in order to increase hybrid fertility had all been performed in *S. cerevisiae* x *S. paradoxus* hybrids. First, a baseline for *S. cerevisiae* x *S. uvarum* spore viability in the lab and

my hands was established by sporulating a hybrid diploid generated from wild type haploids from each species and dissecting progeny. After growth, the survival of these progeny was scored and a percent viability was calculated to be ~0.87%. This value is very similar to previously reported data of ~1% for other inter-species *Saccharomyces* hybrids. After repeating this experiment on the *msh2Δ/msh2Δ* hybrid, I observed this percent viability to increase to around 2.4%. This method of sporulation and dissection was also repeated for the homozygous p_{CLB2}-*MSH2* and p_{CLB2}-*SGS1* hybrid strain as well. The viability of these offspring showed nowhere near the increase seen for the engineered p_{CLB2} *S. cerevisiae* x *S. paradoxus* spores: in fact the viability was much lower than for the *msh2Δ* hybrid strain progeny [Figure 2.2.1.].

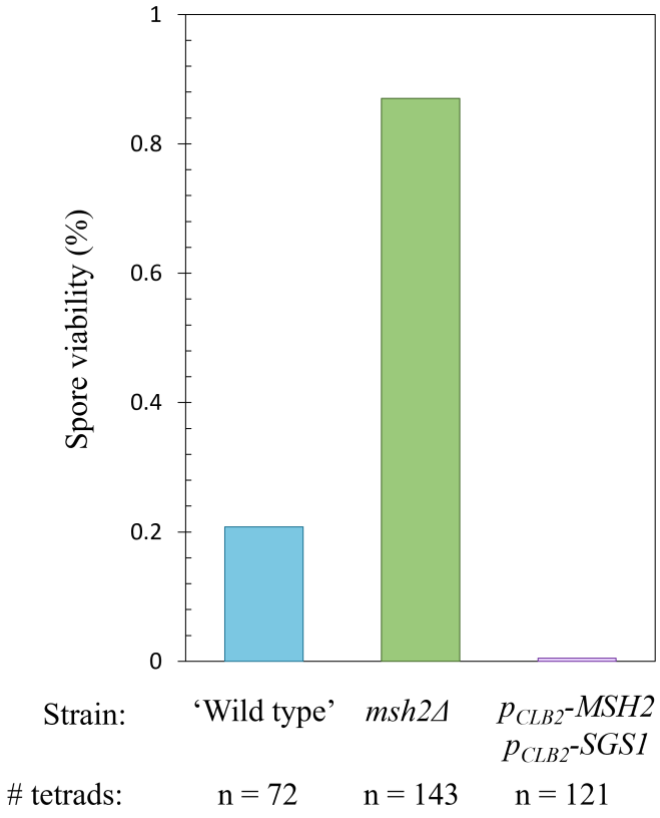


Figure 2.2.1. Comparison of spore viability for diploid hybrid strains. All strains are homozygous at all MMR alleles.

These numbers are much smaller than the ~10% that has been observed for *S. cerevisiae* x *S. paradoxus* hybrids, which share more sequence identity and possess syntenic genomes. Reducing the action of anti-recombination does increase homeologous recombination, but larger chromosome scale translocations as seen in *S. uvarum* when compared to *S. cerevisiae* provide

an additional known barrier to hybrid fertility. This is in addition to other potential factors which may contribute to hybrid fertility, since even chromosomal translocations alone cannot explain the lower hybrid fertility found in *S. cerevisiae* x *S. uvarum* hybrids with mutated MMR machinery[119].

2.2.2 Meiotic null alleles in *S. cerevisiae* x *S. uvarum* hybrids remain mostly infertile but the aneuploidy in recovered offspring is decreased.

In order to analyze further the genomic composition of surviving progeny, an en masse spore retrieval protocol known as random spore analysis (RSA) was utilized. RSA is facilitated by using all parent *S. cerevisiae* strains from the Yeast Deletion Collection. To facilitate more high throughput experiments with this strain collection, an updated version of these strains was engineered to contain a suite of markers collectively known as the Synthetic Genetic Array (SGA). These markers enable efficacious bulk selection of haploid *MATa* progeny from a sporulated diploid. This is done by subjecting the sporulation culture to a combination of mechanical and chemical selections to selectively kill diploids and mechanically separate spores.

One of the key elements to the SGA markers that collectively allow for the selection of haploids includes the *HIS3* gene under the control of a mating type specific promoter. By plating progeny recovered after RSA on media lacking histidine, the only surviving cells will be haploids for only one mating type, which prevents the inter-mating of progeny. In addition, the *CAN1* and *LYP1* genes make it possible to selectively destroy diploids remaining within a sporulation culture. *CAN1* encodes for an arginine permease, a transmembrane protein that facilities arginine uptake

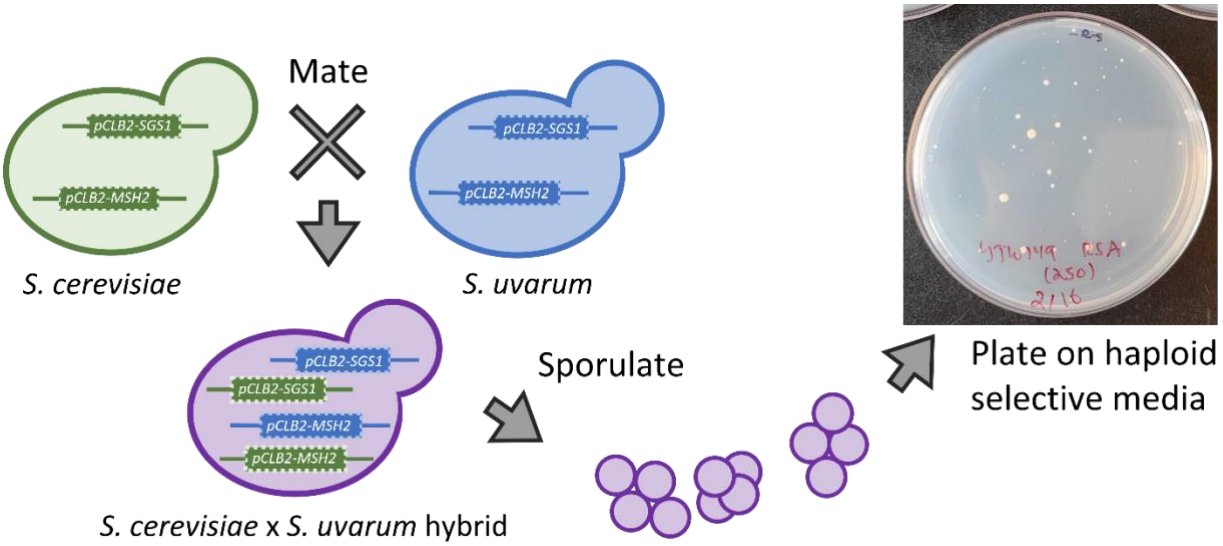


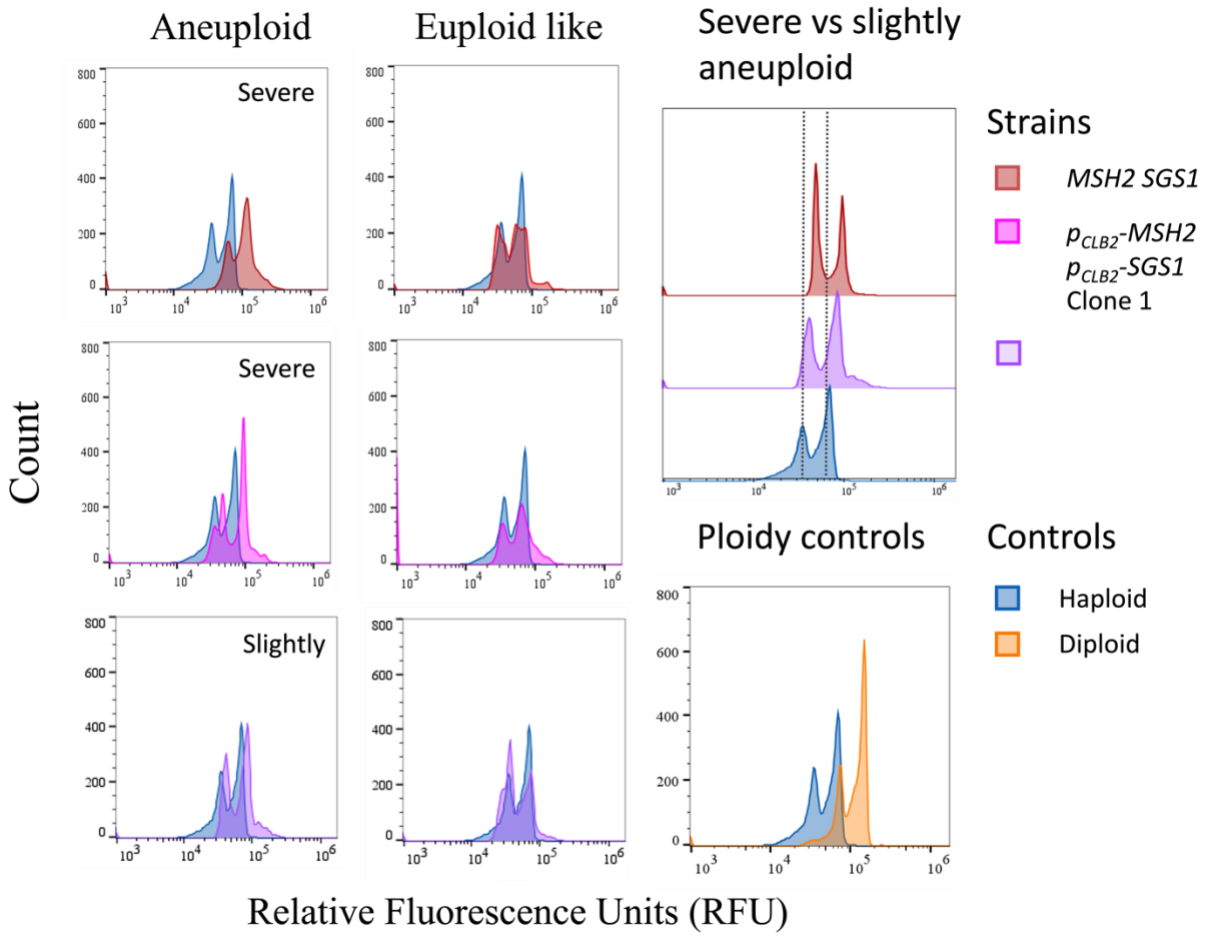
Figure 2.2.3. Diagram of p_{CLB2} promoter replacement hybrid and process of random spore analysis to recover viable progeny.

from the extracellular environment, while *LYP1* encodes for a lysine permease. Both transmembrane proteins are imperfect in target recognition, and are capable of up taking the lethal drugs canavanine and thialysine respectively instead of their correct targets; leading to cell death. By plating mechanically separated and chemically digested spores onto either or both of these drugs, heterozygous diploids are selectively killed [Figure 2.2.3.]. This is because a cell containing even a single functional copy of *CAN1* or *LYP1* is capable of intaking enough toxic analogue to kill the cell. Since haploid offspring should only have one copy of either, progeny that inherit the null alleles will survive on the corresponding drug because there is no permease to uptake the analogue. Leveraging RSA in this way, 96 hybrid offspring for each experimental and control strain were collected and subjected to growth assays and DNA content staining followed by flow cytometry. Analysis was performed in the FlowJo software: comparing cell count by relative fluorescence units histograms for each experimental segregant.

These experimental segregant DNA content staining plots were compared by eye and FlowJo

fluorescence gating pairwise with the results for both a haploid and diploid control [Figure 2.2.4]. The segregants were then binned into several different categories based on the overlap, or lack thereof, of the experimental samples with the haploid control [Table 2.2.1]. These flow cytometry results were then used to calculate a percentage of approximately euploid offspring for each hybrid population: wild type, and two different biological replicates of the homozygous $p_{CLB2\text{-}MSH2}$ and $p_{CLB2\text{-}SGS1}$ engineered strain. As expected, the 'wild type' hybrid strain had the most aneuploid offspring, with the euploid percent at only 7.89%. Compared to this, the

Figure 2.2.4. Comparative DNA staining flow cytometry analysis of hybrid offspring ploidy between wild type and p_{CLB2} MMR meiotic null strains. Representative plots are shown for binning of p_{CLB2} progeny as Severe or Slightly aneuploid. Offset overlay showing in more detail the average difference in aneuploidy levels for p_{CLB2} offspring vs. 'wild type' offspring.



biological replicates for the experimental p_{CLB2} promoter replacement yielded a substantially larger percentage of euploid like offspring at 46.5% and 42.9% respectively. In addition, the aneuploid offspring produced by these p_{CLB2} replacement strains were on average less aneuploid, indicated by the 'Slightly aneuploid' category label.

2.3 Discussion

While the engineered promoter replacement strain may not have yielded an increased percentage of viable spores by manual dissection, the flow cytometry analysis shows that spores recovered by high throughput methods from the engineered strain show a significant increase in euploid like and slightly aneuploid offspring compared to the progeny generated from a 'wild type' hybrid strain. It should be noted that the particular RSA method applied here did yield a rather high percentage of diploid parental strain yeast that sneaked though the selection. I believe this rate of background cells could be ameliorated by leveraging other random spore analysis protocols that subject the samples to a higher frequency of sonication, resulting in the destruction

Table 2.2.1. Summary of hybrid strain information and quantitative binning of flow cytometry ploidy level results for all SGA selected surviving offspring for each parental hybrid strain. Assignment of DNA content stained offspring as Severe aneuploid, Slightly aneuploid, and Euploid was done based on distance of rightward shift of DNA content staining for offspring compared to a haploid control strain.

Strain information		Flow cytometry ploidy binning summary			
Name	MMR genotype	Severe aneuploid	Slightly aneuploid	Euploid like	% Euploid like
YTW125	*MSH2 SGS1*	35	N/A	3	7.89
YTW164	p_{CLB2}-*MSH2* p_{CLB2}-*SGS1* Clone 1	9	14	20	46.5
YTW165	p_{CLB2}-*MSH2* p_{CLB2}-*SGS1* Clone 2	6	17	18	43.9

of cells not protected by the tetrad ascus.

While these results are preliminary and DNA content staining yields no information about recombination rates and crossover events, I conclude that the increased yield of euploid like offspring from a hybrid parent can be greatly increased by the p_{CLB2} promoter replacement engineering method. While the percent euploid like offspring results are very promising, DNA content staining as applied here is a more coarse measurement. It cannot capture recombination, or even if each chromosome is indeed inherited equally, rather than patterns of aneuploidy that mimic a euploid DNA content by DNA staining. Future plans include performing whole genome sequencing on select euploid spores for both the wild type and engineered hybrid strains leveraging next generation sequencing (NGS) technologies. In conclusion, the selective recovery of surviving euploid hybrid progeny by flow cytometry sorting enables the application of genetic mapping methods to infertile *Saccharomyces* hybrids separated by genomic translocations.

2.4 Materials and methods

Strains

For the *msh2* knockout strains, the *S. cerevisiae* haploid was sourced from the *MATa* Yeast Deletion Collection. The creation of the SGA version of this collection was described previously in Tong et al., but in short, each non-essential gene was systematically deleted by a barcoded KanMX cassette within a diploid strain derived from BY4742 already containing an assortment of other markers. Subsequent sporulation of this diploid and selective screening of progeny allowed for recovery of haploid offspring with the SGA marker suite, and the gene of interest deletion cassette. The *msh2Δ* knock out for *S. uvarum* was generated previously by lab member

Monica Sanchez. Briefly, the endogenous genomic locus of *MSH2* was replaced by transforming a wild type *S. uvarum* yeast strain derived from CBS7001 with a PCR product containing the *NatMX* selectable marker flanked by primer designed homology to the endogenous *MSH2* locus. *S. cerevisiae* x *S. uvarum* hybrids were generated by patch mating of parental strains on solid YPD plates and picking zygotes under the microscope after 4-5 hours post mating. A 'wild type' hybrid was generated by mating a haploid version of the SGA strains termed YMD895 to the haploid *S. uvarum* strain termed YTW2.

The haploid *MSH2* and *SGS1 S. cerevisiae* strains used in this work were generated using a diploid *S. cerevisiae* parental strain originating from the synthetic genetic array (SGA) collection originally generated by Tong et al. This strain had the genotype of *can1Δ::STE2-Sp_his5*/*CAN1 lyp1Δ*/*LYP1 his3Δ1*/*his3Δ1 leu2Δ0*/*leu2Δ0 ura3Δ0*/*ura3Δ0 met5Δ0*/ *LYS2*/*LYS2* MET15/met15Δ0 and mas made from BY4742, a direct descendant of the main laboratory *S. cerevisiae* strain S288C. A diploid strain termed YTW3 was used to avoid any suppressor mutations in case of fitness defect in the promoter replacement allele that could occur in a haploid. Each promoter replacement strain was made by transforming the yeast with the appropriate PCR product (plasmid assembly defined below) as described previously[121]. Briefly, a colony of the parent *S. cerevisiae* strain YTW3 was grown overnight. This culture was diluted the next morning and grown until an OD of 0.5 – 0.8 was reached. Cells were then spun down, washed, and subjected to a chemical transformation using lithium acetate, polyethylene glycol, water, and DNA to be transformed. Cells were then recovered for two hours in YPD and finally plated on either D + G418, or D + CloNAT plates as needed. Single colonies were picked to check for promoter replacement via PCR and Sanger sequencing.

S. uvarum strains in this study all originated from the CBS7001 parental strain, whose HO locus has been disrupted to prevent mating type switching. Promoter replacement strains were made similarly to the *S. cerevisiae* strain: transformation of the diploid termed YTW8 with PCR to replace either the *MSH2* or *SGS1* promoter with the *S. uvarum CLB2* promoter marked with a *NatMX* and *KanMX* cassette respectively. The only change in protocol was for the heat shock to occur at 37°C instead of 42°C, and for all normal growth to occur at 25°C instead of 30°C due to the difference in preferred growth temperature for *S. uvarum* compared to *S. cerevisiae*.

Gibson assembly

To replace the native promoters of *SGS1* and *MSH2*, primers were designed to amplify the *CLB2* promoter sequence from genomic DNA of both *S. cerevisiae* and *S. uvarum*. These PCR products were then cloned into the plasmid pFA6a via Gibson assembly using inverse PCR for the plasmid backbone to remove several expression cassettes unnecessary for this experiment and designed overhangs. For each species' promoter, two constructs were created: one with a *KanMX* cassette upstream of the promoter, and the other with a *NatMX* cassette. The *KanMX* cassette was used to demarcate the replacement of the native *SGS1* promoter with p_{CLB2}, and the *NatMX* marker used to indicate promoter replacement for *MSH2* loci in both species.

Spore viability

Hybrid cultures were sporulated in 5 mLs of 1% potassium acetate, 0.05% dextrose, and 0.1% yeast extract for 4-7 days. Ascospore dissection was carried out by spinning down 50 μL of the sporulation culture, resuspending in 15 μL of yeast lytic enzyme (YLE) at 100 U/mL in 1 M

sorbitol, and plating on YPD after a 30 minute digestion at 30°C.

Dissections were carried out at room temperature on a Nikon H550S Eclipse 50i scope via manual needle dissection. After 3 days of growth at 25°C surviving tetrad spores were counted. Percent viability for each hybrid strain was calculated by the following formula:

$\frac{\#\ viable\ spores}{4\ x\ \#\ tetrads}\ x\ 100\%$.

To recover viable progeny en masse in a high throughput setting, the method of random spore analysis was used[122]. Briefly, a volume of sporulation culture for each hybrid was spun down and resuspended in 200 μL of 100 U/mL YLE. After 30 minutes of incubation at 30°C, cultures were subjected to 1 minute of vortexing followed by plating onto C +Canavanine -arginine -serine or C +Thialysine -lysine. Surviving colonies were collected, preserved with a cryoprotective reagent, and subsequently subjected to growth analyses and flow cytometry

Flow cytometry

In order to determine the level of aneuploidy among hybrid spores, DNA content flow cytometry was performed using a BD Accuri C6 flow cytometer. The protocol is as described previously in the 2015 edition of Methods in Yeast Genetics and Genomics (A Cold Spring Harbor Laboratory Course Manual). Briefly, spores as well as haploid and diploid controls were seeded to grow overnight at the appropriate temperatures. Cultures were back diluted to ~0.2 OD the next day and then harvested between ~ 0.5 – 0.8 OD by spinning down 1 mL and resuspending in 1 mL of 70% ethanol. Half of these fixed cells were then washed with water and subjected to RNase A digestion in 200 μL of 50 mM Tris-HCl (pH = 8.0) for 2 hours at 37°C. Samples were then spun down and resuspended with 2 mg/mL Proteinase K in 200 μL of 50 mM Tris-HCl (pH = 7.5) for 30 minutes at 50°C. After another spin, samples were resuspended in 200 μL of FACs buffer

(200 mM Tris-HCl pH = 7.5, 200 mM NaCl, and 78 mM $MgCl_2$). Directly before samples were loaded onto the cytometer, 10 μL of cells were transferred to a 96-well plate containing 300 μL of SYBR green diluted 5000X in 50 mM Tris-HCl (pH = 7.5) and subjected to sonication. All samples were sonicated for 3 seconds at 10% power, resulting in a 5-10 Watt power output. Analysis of flow cytometry samples, including gating for debris and doublets, was carried out in FlowJo.

Chapter Three: Rescue-by-mating as a method to facilitate strain engineering with haploid *Saccharomyces* strains containing essential gene knockouts.

3.1 Introduction

3.1.1 Previous work described rescue by mating for intra-species *S. cerevisiae* crosses.

Discussion so far in this thesis has focused on more traditional methods for genetic mapping approaches that can be leveraged to map the interacting rescue loci underlying differential gene essentiality between *S. cerevisiae* and *S. uvarum*. However, there do exist other methods that can be leveraged as mapping approaches that have not been discussed yet. One such approach that requires no extensive strain engineering because it does not require infertile hybrids to produce viable progeny is transposon mutagenesis. However, the particular hybrid strains necessary for the version of transposon mutagenesis screening of interest are non-trivial to engineer. There will be an extensive discussion of transposon mutagenesis approaches later in this thesis, but this chapter will focus on a method to generate the strains required for such an approach.

In order for a transposon mediated mutagenesis screen to work for mapping the interacting loci underlying differentially essential genes, the hybrid must be homozygous knockout for both orthologs of the gene of interest. In this strain, transposon mutagenesis is leveraged to perform a specific genetic screen called reciprocal hemizygosity. A reciprocal hemizygosity screen is generally used to discern phenotypic differences in a hybrid by alternatively disrupting function of one parental allele or the other. The loci within the hybrid genome that are responsible for the differential gene essentiality phenotype can be mapped with this method. For each candidate gene that is differentially essential, an additional gene knockout due to transposition in a

homozygous knockout hybrid strain will have a different effect depending on if the *S. uvarum* or *S. cerevisiae* parental allele is interrupted by transposition. In order for this particular screening approach to work, both parental alleles of the differentially essential gene must already be disrupted before transposition so that neither parental allele is present as a 'wild type' sequence that would prevent the mapping of other interactors.

The experimental set up for such an approach is significantly different than if genetic mapping is done in the style of QTL. The initial experimental strain is constructed by mating corresponding haploid *S. cerevisiae* and *S. uvarum* strains that each bear the deletion of the species specific ortholog for a differentially essential gene. This process generates a diploid *S. cerevisiae* x *S. uvarum* hybrid that is homozygous knockout for both orthologs of a differentially essential candidate gene. Specifically, the work in this thesis focuses exclusively on differentially essential genes that are essential in *S. cerevisiae*, but nonessential in *S. uvarum*. This means that generating haploid *S. uvarum* mating partners that have the candidate differentially essential genes knocked out is relatively simple. However, a problem arises in the need to generate the corresponding *S. cerevisiae* mating partner strain with its own ortholog deleted. Because these differentially essential gene candidates are essential in *S. cerevisiae*, any attempts to generate a haploid deletion strain like for the corresponding *S. uvarum* nonessential ortholog will fail.

Observed in the lab as a byproduct of the planned experiment, rescue-by-mating has now fully become a field of research interest as well as a helpful tool for strain engineering—of particular interest to us[123]. Rescue-by-mating describes the phenomenon by which a haploid spore generated from meiosis of a diploid parental strain can be rescued from cell death by mating to a

viable spore, often from the same ascus in nature, to produce a viable diploid[124]. This rescue of inviable haploid progeny was of particular interest to our work because of the inherent challenges present in attempting to generate hybrids homozygous for LOF alleles of a differentially essential gene. Specifically, the challenge present in trying to generate haploid knockouts for essential genes necessary to form a hybrid that is homozygous knockout for a differentially essential gene.

While there are some clever strain engineering tricks leveraging *Saccharomyces* resources that can be utilized, rescue-by-mating offered a viable alternative. In this method, spores produced by an intraspecies diploid parent heterozygous for the essential gene deletion would be mixed for mating with a haploid strain from the other species that already has the differentially essential gene ortholog deleted. The existence of this haploid strain would be possible since it would be nonessential in that species. After this mating, selection and screen would yield hybrid diploids that are homozygous for the differentially essential gene deletion because mating would occur before the spore with the essential gene deletion underwent cell death. These strains would be critical to generate because they would provide the background for the transposon mediated genetic mapping of the interacting loci for the differentially essential gene candidates.

3.1.2 Potential application of rescue by mating to work with haploid essential gene knockout strains.

The method of rescue by mating offers a strategy to work with haploid knockout strains of essential genes. Previous work has proved that during the process of spore generation, it is possible for the mother cell to form four haploid spores that each could be lacking essential

genes[123,124]. These spores remain dormant within the ascus as long as nutrients are still in scarce supply, in starvation conditions. After encountering rich media, the spores will be triggered to begin germination. This is where the inviability of haploid essential gene knockout strains becomes apparent very quickly. If one were to dissect a heterozygous knockout strain for an essential gene on rich media, after growth two of those original spores would form full colonies, while two would not grow.

However, if one were to examine these tetrads under the microscope immediately after dissection, one would be able to see that for some essential genes the haploid knockout spores will have undergone several cell divisions to form a microcolony before the cells become incapable of further mitotic growth and undergo cell death. Previous work with *S. cerevisiae* diploids heterozygous for a lethal mutation from intra-species matings has shown that in this time period, directly after germination, a spore that will become inviable can actually mate with an available partner of the opposite mating type during its few, viable cell divisions. With this logic, I planned to attempt rescue by mating between haploid essential gene deletion *S. cerevisiae* strains and a *S. uvarum* mating type partner. If this method worked, it would prove that rescue by mating works even across species divides within the same clade, in addition to providing the utility of a way to make the challenging homozygous knockout hybrids strains necessary for transposon mutagenesis.

3.2 Results

3.2.1 Rescue by mating functions across species barriers with *S. cerevisiae* x *S. uvarum* crosses.

Rescue by mating studies were carried out with substantial effort of another graduate student in the lab, Abigail Keller. We followed previous mass haploid progeny recovery methods developed to selectively destroy any remaining unsporulated diploid cells in a sporulation culture, as well as removing the ascus surrounding the spores. Sporulation cultures of heterozygous essential gene knockouts were treated with an array of chemical and mechanical stresses as described in the Methods. After these treatments, individual spores were recovered from the original sporulation culture for each heterozygous essential gene knockout strain. The predicted differentially essential candidates studied here included *AME1*, *BRR6*, *ORC6*, and *SPT5*[55]. These candidate genes represent a wide array of cellular biological processes, and whose participation in these different functions is suggestive of potential interacting partners that could include the causal loci for differential gene essentiality. The identity of these specific strains from the *S. cerevisiae* heterozygous deletion collection was verified by PCR and sequencing of the barcodes from the original deletion collection construction.

Table 3.2.1 Rescue by mating results for candidate *S. cerevisiae* species specific essential genes. Microcolony formation was observed under a dissecting microscope.

***S. cerevisiae* specific essential gene candidates**	**Microcolony formation**	**Pre-mating spores OD**	**Recovered hybrid diploid colonies from 1:10 dilution plating**
AME1	2-3 divisions	0.33	7
BRR6	3-4 divisions	0.36	2
ORC6	1-2 divisions	0.43	3

The post treatment spores for each candidate gene were then split into two pools. One of these pools was mixed with a MAT*a* 'wild type' *S. uvarum* partner, while the other was mated to a MATα *S. uvarum* partner. This is in order to recover as many rescue by mating events as possible: whether the temporarily viable spore with an essential gene knockout inherited MAT*a* or MATα. These mixed matings were incubated overnight and then surviving cells were plated

onto selection for the *KanMX* cassette marking the essential gene deletion and a functional *URA3* that could only originate from the *S. uvarum* genome, resulting in varying numbers of recovered hybrid diploid colonies [Table 3.2.1]. Surviving colonies were verified to be *S. cerevisiae* x *S. uvarum* hybrids by sequencing of a target locus. We found that rescue by mating between *S. cerevisiae* and *S. uvarum* partners was possible but occurred at quite a low frequency. While reasonable cultures sizes of 5 mL were used in this study, it would potentially be possible to scale up to larger volumes if it was necessary to recover more mated hybrids. If the percentage of hybrid diploids recovered from rescue by mating remains constant, by scaling to larger experimental volumes one can recover more of these rescue by mating events for further study or use in transposon mutagenesis experiments.

3.3 Discussion

Previous work studying the control features and processes of spore generation and initial spore germination has provided a wealth of information about the biology of *S. cerevisiae* spore generation. An added, and perhaps unexpected, benefit of this knowledge is that *S. cerevisiae* spores that lack any of the functional genes required for viability are in fact potentially able to begin germination. These spores can also even undergo several cell divisions before undergoing cell death triggered by the catastrophic cellular consequences of missing any of the approximately ~1000 essential genes in *S. cerevisiae*. The power in this knowledge also lies in expanding the tool kit for scientific groups interested in studying these essential genes and their viability.

This finding was even more interesting for the work discussed here, because of general

difficulty to engineer the strains necessary for transposon mutagenesis mapping of the genetic architecture of differential gene essentiality between *S. cerevisiae* and *S. uvarum*. While there exist other tools in the arsenal to work with essential genes for the homozygous knockout hybrid strain generation, rescue by mating offered another method by which to create these strains necessary for the transposon mutagenesis method to be used. Rather than working with any exogenous complementation systems to rescue a haploid essential gene knockout *S. cerevisiae* strain, rescue by mating offered a viable opportunity to avoid plasmid complementation completely. While plasmid complementation for strain engineering presents a viable method, the experimental pipeline to generate homozygous knockout hybrids this way is longer and more involved. This method of complementation also requires more consecutive steps, which each take a nonsignificant amount of time.

Instead, rescue by mating made it possible to leverage pre-existing resources within the community such as the heterozygous gene deletion collection with the SGA marker suite generated by Tong et al. The candidate differentially essential genes—*AME1*, *BRR6*, *ORC6*, *SPT5*—could be selected from the overall gene deletion collection, verified by barcode sequencing, and sporulated. By leveraging the rescue by mating protocol and mixing with an *S. uvarum* mating type partner, we showed it was possible to obtain diploid hybrids. Crucially, rescue by mating made it possible to generate a *S. cerevisiae* x *S. uvarum* hybrid where the *S. cerevisiae* mating partner had a nonfunction allele of an essential gene.

This method opens the door to facilitating larger scale generation of *S. cerevisiae* x *S. uvarum* hybrids where either parental alleles of the differentially essential candidate genes are knocked

out. Rather than selecting only a few candidate genes of interest, leveraging the heterozygous deletion collection in *S. cerevisiae* and even some preliminary, albeit still incomplete, *S. uvarum* deletion collections enables larger scale screening. High throughput screens make it possible to interrogate the genetic architecture underlying these differentially essential genes between *S. cerevisiae* and *S. uvarum* more deeply. In addition, the novel finding that rescue by mating within hybrids works also points towards interesting biology perhaps of the level of conservation between the rescue by mating machinery across the two species within the *Saccharomyces* clade. Even with the power of this method, it is important to keep in mind that the success of rescue by mating can be dependent on the identity of the gene with a lethal mutation: as any involved in early germination necessary for rescue by mating would be unable to undergo the few cell divisions within which mating occurs.

3.4 Materials and methods

Strains

All four of the candidate differentially essential *S. cerevisiae* strains were sourced from the heterozygous deletion collection originally constructed by Tong et al. as described there, which then had the SGA marker suite engineered later[125]. Their identity was each confirmed by PCR and Sanger sequencing of the uptag barcodes from the deletion collection. The *S. uvarum* rescue mating partners were both derived from the CBS7001 strain background.

Sporulation

A single colony of each heterozygous essential gene knockout *S. cerevisiae* strain was inoculated overnight in 3 mLs of YPD. The next day, cultures were back diluted to ~0.2 OD in fresh YPD

and outgrown for two more hours. 1 mL of each of these cultures was spun down and washed twice with 1 mL of ddH_20. These washed cultures were each resuspended in 5 mL sporulation media consisting of 1% potassium acetate, 0.05% dextrose, and 0.1% yeast extract and placed in a roller drum at 30°C for 4-7 days.

Rescue by mating

After sporulation was confirmed, 1 mL of the culture was centrifuged for 10 seconds at 13,000 xg. The resulting pellet was resuspended in 5 mL of ddH_20. To this suspension 100 μL of 100 Unit/mL YLE, and 10 μL of 2-mercaptoethanol were added. These tubes were incubated overnight at 30°C with gentle shaking. The next day 5 mLs of 1.5% triton X-100 was added and the tubes were then incubated on ice for 15 minutes. Sonication was performed with the microtip function of a MISONIX Ultrasonic Liquid Processor. The sonicator probe was inserted into the 10 mL of spores within each tube such that it did not touch the sides or bottom and sonication performed as followed: microtip power setting Amplitude of 30, with three cycles total of 20 seconds of sonication and 1 minute on ice. This was repeated for each sample, with the microtip being cleansed thoroughly between samples with 70% ethanol. These sonicated spores were then spun down for 10 minutes at 1200xg. The resulting pellet was resuspended in 5 mL of 1.5% triton X-100 and vortexed vigorously. These spores were then spun down again for 10 minutes at 1200xg and washed once more. After the final centrifugation the spores were then resuspended in 5 mLs of ddH_20 and a 10-fold dilution counted by hemacytometer to confirm the absence of any ascus or spore clumps. Spores were then diluted if necessary to 1 $x10^3$ spores/mL. Matings with the *S. uvarum* rescue partner were carried out by mixing 200 μL of spores with 200 μL of log phase *S. uvarum* haploid mating partner at approximately the same density as the spores. A

final 200 μL of YPD media was added to this mating mixture, and the mixture of spores and S. uvarum haploid mating partner was incubated in a roller drum overnight at 30°C. The following day dilutions of the mating mixtures were plated on C -Ura +G418 to select for diploids carrying the G418 deletion allele for the *S. cerevisiae* essential genes and a functional *URA3* allele that can only contributed by *S. uvarum* mating partner genome.

Chapter Four: Determining the genetic basis for differential gene essentiality in two *Saccharomyces* yeasts through transposon mutagenesis mapping.

4.1 Introduction

4.1.2 Motivation and history of mapping genotype to phenotype

For over a century, the fields of biology and genetics have been focused on trying to understand genotype to phenotype connections in the natural world. Whatever driving question is being explored, the field remains unified in the central goal of building a general understanding for the phenotypic consequences of mutations, whether de novo or inherited. Building a predictive understanding of the phenotypic consequences for genomic mutations is especially important for applications in human health and disease. Such work would also further the field of evolution's understanding of how standing variation and accumulated mutations are acted upon by natural selection within populations of organisms.

There exists an array of methods that have been applied to this problem with aplomb: generating rich datasets on a candidate gene by candidate gene basis. These bodies of work usually focus their efforts within a single genetic background of whichever model organism they are working with[7,45]. Over the years, the understanding of the field for the effect of genetic background—the standing genetic variation that exists within an individual of a species that can interact with novel mutations in the gene of interest and thus confound genotype to phenotype studies of a gene of interest—has evolved. Once considered a nuisance confounding the original studies, genetic background effects have now become a field of study in their own right, part of a scientific pilgrimage to plunge deeper yet into understanding the biology of genomes and the connections between genotype to phenotype. It should be noted that single genetic background DMS on a

gene of interest data sets are still a rich wealth of information for mapping the sequence to function relationships for genes of interest and have even been applied within human health to investigate the phenotypic consequences of rare or common variants in genes like *BRCA1* or *CYP2C9* to empower doctors and patients to provide targeted care for their patients: a field known as precision medicine[126–128].

4.1.2 Genetic mapping in the endogenous context of a gene within its genetic network.

Even with the power of en mass saturated mutagenesis studies, an important drawback of such approaches is that they treat genes as islands within a cell, such that only variants in the gene of interest will have an effect on the phenotype of interest. This is a simplification of the endogenous context for these genes within the cell, where their products interact either directly or indirectly with other gene products in the cell to perform their functions. These interactions are described as genetic network interactions, or epistatic interactions. They are defined by comparing individual fitness effects for loss of function mutations for two (or in some cases more) different genes to the phenotypic fitness result of all of these mutations within the same organism[18,32,129–132].

The expectation for two or more genes not experiencing a genetic interaction is the individual fitness scores will show a multiplicative effect when compared to the combinatorial loss of function strain[30,38,130]. E.g. if the fitness score, on a scale from 0 to 1, for gene A is 0.5 and the score for a gene B loss of function is 0.2, in a double knockout strain without gene A or gene B the overall fitness score would be 0.1. If the combined fitness score deviates from this number, a genetic interaction is assigned between gene A and gene B: if the number is smaller than 0.1 it

would be termed a negative epistatic interaction, and if the number is greater than 0.1 it would be assigned as a positive epistatic interaction[17,21,131,133,134]. While there are no hard and fast rules for predicting genetic interactions based on functional information alone, there do exist some patterns observed for the molecular mechanism underlying epistasis. Generally, genes in negative epistatic genetic interactions tend to function in redundant or overlapping pathways or even as subunits within the same larger protein. While genes in a positive epistatic interaction often are members of the same biological pathway, with one gene downstream or upstream of the other within the pathway. A genetic interaction does not always predict a physical protein-protein interaction, and these interactions are often interrogated differently[22–24,29,135]. It has also been observed that genes that function within the same local network tend to share similar expression patterns, and concerted changes in expression under different conditions.

Mapping these endogenous gene by gene interactions has not been an undertaking for the faint of heart: due to differences in ploidy and level of genetic tractability as well as the availability and efficacy of genetic tools, mapping these interactions can vary by method from organism to organism[16,18,19,21,34,43,59,125]. For example, the most complete genetic interaction network to date exists within *S. cerevisiae*, where the method of screening gene by gene interactions can be automated thanks to clever strain engineering and the genetic tractability of *S. cerevisiae* in general. Due to its high rate of homologous recombination, it is possible to generate single gene deletion strains through mass transformation one by one of *S. cerevisiae* with a barcoded PCR product where regions homologous to the target gene flank a dominant drug marker selection cassette conferring G418 resistance known as *KanMX*. After strain engineering, an automated platform can be used where each well contains a final strain that possesses a target gene deletion

whereby the endogenous gene is replaced by this *KanMX* cassette either in a haploid or as a heterozygous diploid. Later work would additionally add another suite of markers to this collection of strains known as the Synthetic Genetic Array (SGA). These markers altogether provide a way to select for a specific mating type in a haploid spore, as well as the ability to selectively destroy parental diploid cells in order to recover only haploid progeny from a sporulation culture.

The Boeke group then built upon this powerful collection by building methods collectively known as diploid-based synthetic lethality analysis by microarray (dSLAM). dSLAM makes it possible to transform the heterozygous gene knockout collection with a deletion cassette for a gene of interest in order to generate a heterozygous knockout of a gene of interest in each strain of the knockout collction. After generation of the knockout, mass sporulation and selective plating is used to specifically recover "one by all" haploid double knockouts containing the original knockout as well as another novel deletion[130]. This method was used to map the genetic interaction networks in *S. cerevisiae*. While the profiling of the single deletion collection discussed previously was utilized to define a list of essential genes for budding yeast. In fact it is from the essential gene dataset for *S. cerevisiae* that a previous graduate student in the Dunham lab originally compared their own mapping method in *S. cerevisiae* to, and then compared this list to *S. uvarum* orthologs to define differentially essential genes.

Although SGA and dSLAM provide excellent methodologies to map essential genes and genetic interactions, the previous lab member was interested in profiling essential genes in *S. uvarum*, which did not have a comparative deletion strain library constructed. Attempting to create this

resource for *S. uvarum* would have been the work of several graduate students' degrees and thus did not present a viable option to perform this study. Thus, the graduate student turned to another powerful genetic method previously used to profile essential genes: transposon mutagenesis.

4.1.3 A brief history of transposons and their application as genetic tools

Transposons are naturally occurring selfish genetic elements first described in the 1940s by Barbara McClintock and her graduate student Harriet Creighton. Since their discovery, an entire field of research has focused on their biology and behavior in nature as well as their potential to be co-opted by geneticists and wielded as a tool for genetic studies. By isolating the sequence of a variety of Class II transposons—transposons that function via a cut and paste mechanism and encode for the transposase necessary to perform this function—geneticists were able to alter these systems[136–138]. Usually, a functional transposon mutagenesis system has separated these sequences such that the flanking arms recognized by the transposase as the DNA to cut and paste now flank a payload such as a drug resistance marker. While the transposon system can function *in vivo* or *in vitro*, usually the transposon payload and transposase are cloned onto an exogenous, bacterially derived expression vector. During the mutagenesis experiment, the transposase can be expressed constitutively or inducibly depending on the selection of the promoter from a specific organism depending on the application.

The very first examples of transposon mutagenesis systems applied to profiling the essentialome for a species occurred in the late 90s and early 2000s after the sequencing of smaller genomes such as viruses and bacteria made it possible to have prior knowledge necessary for the mapping of insertion sites to genomic loci. One of the very first studies of transposon mutagenesis

mapping of essential genes was done by Lynn Enquist and Gregory Smith[139]. They cloned the genome of the pseudorabies virus (PRV) into a plasmid named pBecker1 capable of being maintained in *E. coli*. *E. coli* cells already transformed with pBecker1 were then also transformed with a plasmid containing a mini version of a common transposon system known as Tn5 that is widely used today. This second plasmid also contained the payload encoding kanamycin drug resistance flanked by the transposon arm sequences recognized by the transposase. This transposition occurred within an *E. coli* population *in vivo* en masse such that each cell in the population experienced a transposition event where the kanamycin transposon payload was excised from its original plasmid and inserted into the pBecker1 viral genome. These viral genomes were then extracted from bacteria and used to infect both animals and human cell lines, after which they were then extracted for sequencing of the viral genomes. By calculating the infectivity levels of these transposon mutagenized viral genomes, the group was able to draw conclusions about the viral genes essential for infectivity.

Other transposon mutagenesis experiments since have been carried out in similar manners, some studies such as ones profiling essential genes in *E. coli* itself have followed in the footsteps of the viral study, whereby a mini transposon is leveraged for *in vivo* mutagenesis within the context of the cell. An example of a transposon mutagenesis system used in eukaryotes is SAturated Transposon Analysis in Yeast (SATAY),which enabled one step mapping of essential genes as well as genetic interactions and libraries generated by this method can then be interrogated under an array of experimental conditions to further reveal condition specific genetic effects[140]. The system in SATAY is an *in vivo* one, built on the original transposon system discovered by McClintock and Creighton: the Ac/Dcs system from Maize.

The original strain in which transposition will be performed has a complete deletion of the endogenous *ADE2* locus. The strain is transformed with a plasmid containing a nonfunctional *ade2* gene disrupted by the MiniDs transposon so that cells pre-transposition will not be able to grow on media lacking adenine [Figure 4.1.1.]. A hyperactive form of the transposase AC is inducibly expressed from the same plasmid when the cells are grown on media containing

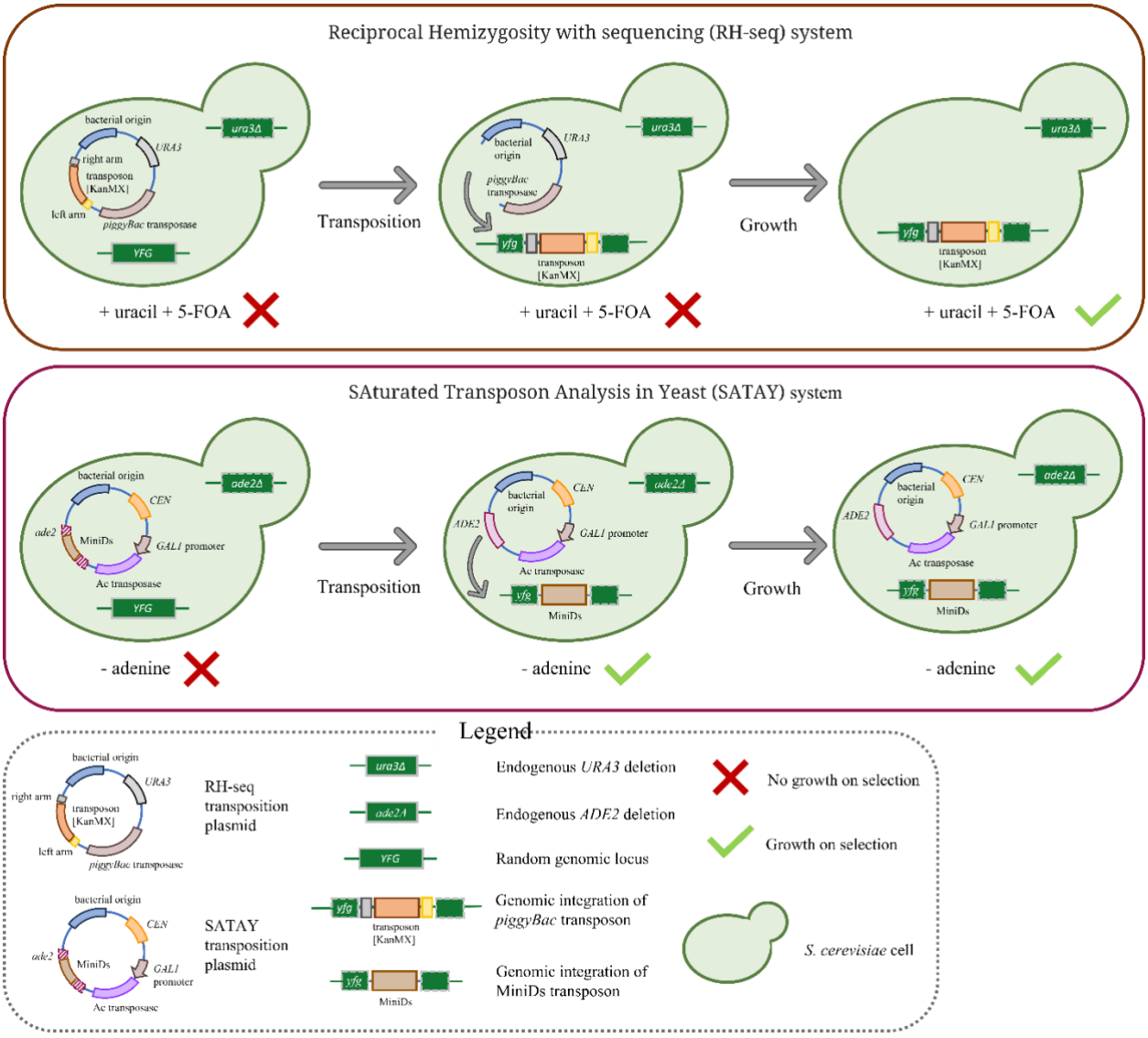


Figure 4.1.1. Schematic diagrams depicting the differences between SATAY and RH-seq as transposon mutagenesis approaches in *S. cerevisiae*. Both depictions begin with *S. cerevisiae* cells which have already been transformed with the respective transposition plasmid for RH-seq or SATAY.

galactose because the native promoter of the transposase has been replaced with a *S. cerevisiae GAL* gene promoter that is turned off in glucose or other sugar sources but switched on in galactose. After induction of transposase expression, the cells are plated on media lacking adenine such that any surviving colonies are those that have undergone transposition, where the MiniDs transposon has been excised by the inducibly expressed AC transposase, resulting in a now repaired and functional *ADE2* gene on the plasmid. Any cells that do not undergo transposition are unable to produce their own adenine and will not be able to grow on adenine drop out media. The surviving cells which can grow on – adenine media because they have undergone transposition can then be pooled and DNA extracted and subjected to transposon enrichment sequencing to map locations of insertions. A lack of insertions is the common signature used to identify essential genes, because any cells that undergo transposition into an essential gene will therefore not be able to survive. While this method is extremely powerful, there are other ways to perform transposon mutagenesis in *S. cerevisiae* and other organisms as well.

It is also possible to perform *in vitro* mutagenesis with transposition, whereby purified genomic DNA from the organism of interest is subjected to random shearing and cloned into a plasmid to create libraries of genomic fragments[55]. These genomic fragment plasmids can then be subjected to transposon mutagenesis within a cell free extract. In the specific case of how the lab member leveraged *in vitro* transposon mutagenesis to profile essential genes in *S. uvarum*, these transposed genomic DNA fragments were then excised from the plasmid, and either the *S. uvarum* haploid and diploid strains were transformed with this library of fragments in order to replace endogenous genes with their transposon disrupted alleles. With this library of

approximately 50,000 unique insertions, by comparing the haploid and diploid insertion sites the lab member was able to create predictive profiles for essential genes, as an essential gene should survive transposition in the diploid, but not the haploid. With these identified essential genes, the lab member was able to generate her lists of differentially essential genes. This list was the resource I utilized to select the differential essential gene candidates I was interested in mapping the underlying genetic architecture of for my thesis work. While this method was powerful enough to answer the lab member's questions, when beginning my plans to leverage transposon mutagenesis as a method to map the interacting loci involved in differential gene essentiality, I turned to more recent transposon mutagenesis methods.

4.1.4 The application of a recent transposon method termed RH-seq to genetic mapping in sterile *Saccharomyces* hybrids.

The primary reason to turn to transposon mutagenesis as a genetic mapping method for my thesis stems from the inability to apply classical QTL mapping techniques in *Saccharomyces* inter-species hybrids. This is because these hybrids are sterile, and thus incapable of producing viable progeny that would be required for QTL mapping. Thus, a workaround for this barrier is performing the mapping within the parental hybrid itself, and therefore avoid any need for the production of viable progeny. The potential pitfall of turning to transposon mutagenesis is that most previous studies applying this method in yeast have not focused on hybrids, where a question mark for the potential efficacy of the transposition within a interspecies *Saccharomyces* hybrid genome exists. Fortunately, I was able to build upon prior work from the lab of Rachel Brem whereby they adapted a piggyBac transposon system for use in *S. cerevisiae* x *S. paradoxus* hybrids[132,138,141].

With a goal to map the underlying genetic architecture for the differing thermotolerance between *S. cerevisiae* and *S. paradoxus*, the Brem group built upon a previous plasmid borne transposon system. While the pre-integration of the transposon in the SATAY method is powerful, issues do still remain for efficiency of the cut and paste within the genome stemming from efficacy issues of the inducible transposase expression. Another drawback of the original MiniDs system is a significant false positive rate whereby colonies that did not actually undergo transposition still grow on the adenine drop out media. As a workaround, the Brem lab adapted a plasmid borne system where the *KanMX* cassette functions as the transposon payload, and the *piggyBac* transposase is constitutively expressed from the same plasmid without the need for induction. This plasmid is cleverly designed to not be maintained by the cell after transposition because of a counter selectable marker, and the lack of a yeast centromeric sequence on the plasmid.

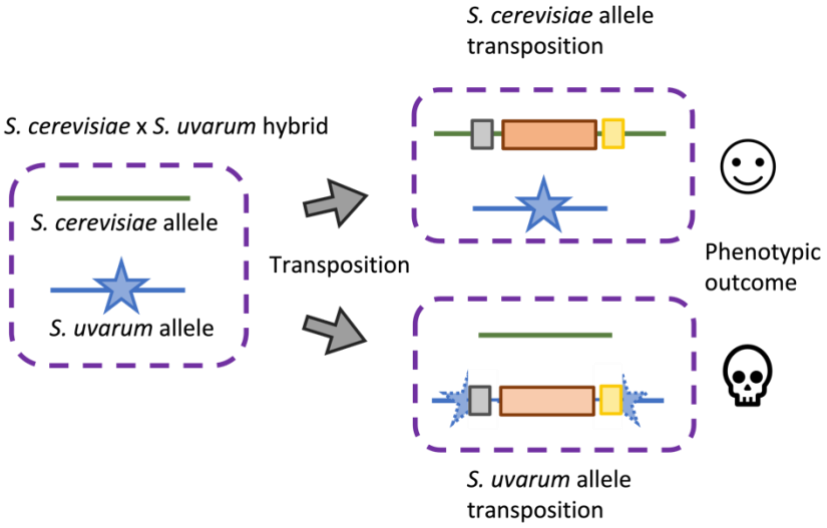


Figure 4.1.2. General depiction of how a reciprocal hemizygosity approach provides information about the species specific interacting factors underlying differential gene essentiality between *S. cerevisiae* and *S. uvarum*.

Armed with this system, the Brem lab was able to interrogate the genetic basis for thermotolerance differences between *S. cerevisiae* and *S. paradoxus* by generating transposon mutagenesis libraries where each cell in a transformation population

undergoes a separate transposition into the genome, disrupting the function of whatever gene the transposon lands within. By assaying the growth temperature preference of these populations and leveraging transposon enrichment sequencing, the Brem lab was able to map the genetic architecture underlying thermotolerance based on the phenotypic consequences for transposon disruption of the genes important for each species preferred growth temperature. The efficacy of the transposon system, the ability to reduce false positives of transposition from generated mutagenesis, libraries, as well as the ability of the system to map more complex underlying genetic architectures of the trait of interest all make this system an ideal method for my own applications. By leveraging this method to interrogate differential gene essentiality within *S. cerevisiae* x *S. uvarum* hybrids, I should be able to determine the genetic interacting partners for our differentially essential candidate genes based on transposon insertion rates that display parental biases specifically to the non-essential species' genome. Technically termed reciprocal hemizygosity screen, this method means that we can ask how a transposon LOF insertion in either the *S. cerevisiae* or *S. uvarum* parental genome present in the hybrid affects the viability of the resulting mutant strains [Figure 4.1.2]. Specifically, this method should allow us to map the interacting factors within *S. uvarum* that suppress the essentiality of its ortholog, which is essential in *S. cerevisiae.*

4.2 Results

4.2.1 Pilot studies of RH-seq transposition within *S. cerevisiae* x *S. uvarum* hybrids proves efficacy of method.

Before fully committing to transposon mutagenesis as my genetic mapping tool for differential gene essentiality, I first wanted to check the efficacy of transposition within a *S. cerevisiae* x *S.*

uvarum hybrid. While it was known at this point in time that the system was capable of working within the *Saccharomyces* interspecies hybrid of *S. cerevisiae* x *S. paradoxus*, the question remained as to whether the system would also work in *S. cerevisiae* x *S. uvarum* hybrids. To test the functionality of the transposon system, I would perform transposon mutagenesis in a 'wild type' *S. cerevisiae* x *S. uvarum* hybrid as well as an intraspecies *S. cerevisiae* diploid. Performing the transposition in the *S. cerevisiae* diploid was a control to make sure the system worked at all in my hands within the setting of our lab and our available resources.

I carried out transposition in both of these strains as described previously to generate pools of transposon mutagenized strains for both the 'wild type' hybrid and *S. cerevisiae* diploid. Briefly, log phase cells were transformed with the plasmid containing the piggyBac transposase and *KanMX* transposon payload. After several series of back dilution and outgrowth to increase loss of plasmid post-transposition, populations were plated on transposon selection, and plasmid counterselection media of C + 5FOA +G418 +URA and then scraped and pooled to generate the final transposon mutagenesis libraries. These libraries, as mentioned previously, are actually populations of cells where each cell in the original transformation culture underwent transposition at a different locus in the genome [Figure 4.2.1].

To test whether this transposition actually occurred, eight single colonies from both the hybrid and *S. cerevisiae* diploid transposon mutagenesis populations respectively were isolated and then subjected to further genetic tests to discern whether transposition had occurred, and whether different transpositions sites within both parental genomes of the hybrid were observed. To discern the first question, I leveraged the single colonies isolated from the *S. cerevisiae* diploid.

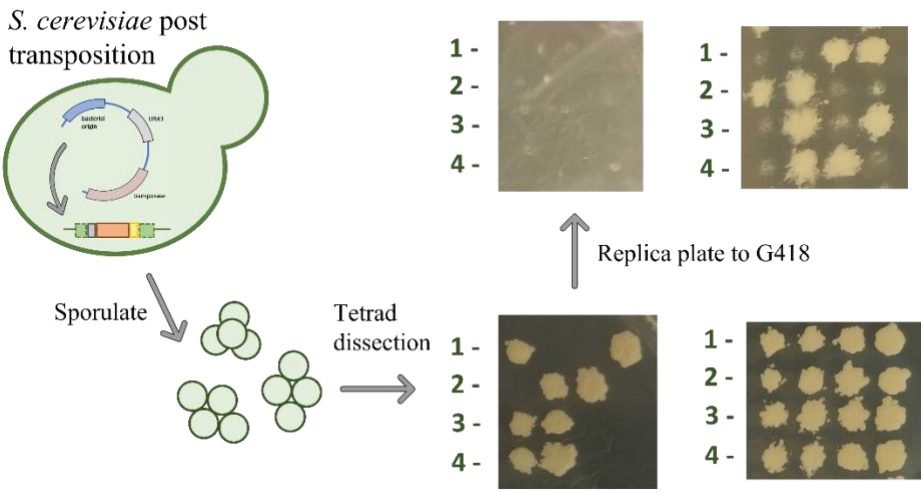


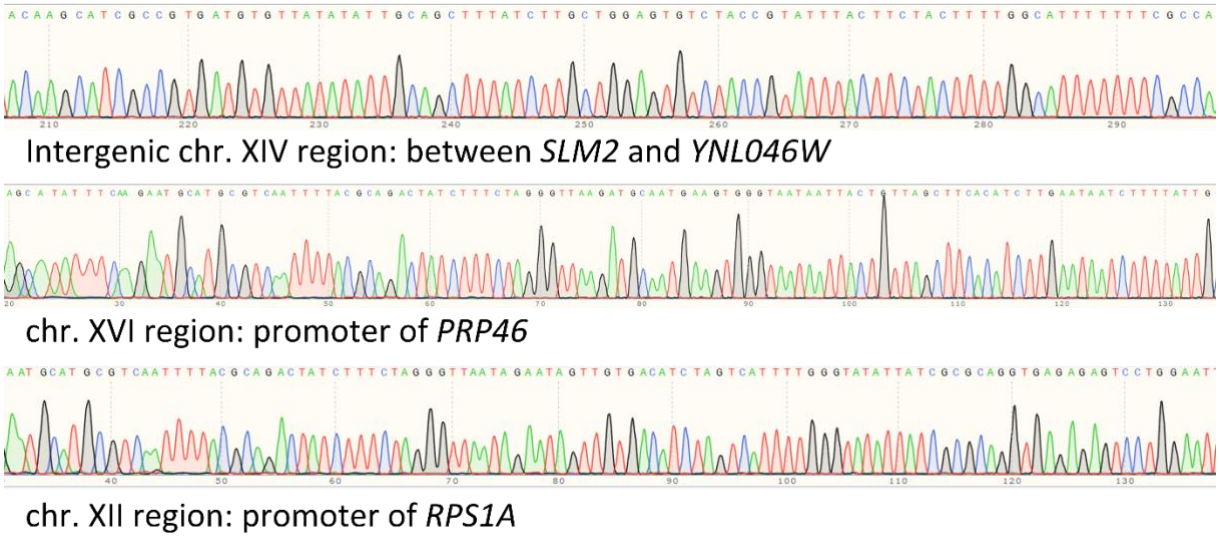
Figure 4.2.1. Post transposition *S. cerevisiae* colonies were sporulated and tetrads were dissected. Two example dissected colonies of transposition into a (left) essential and (right) nonessential gene on YPD and post replica plated onto YPD + G418.

Since the transposition event occurred once within the cell, the genomic locus at which it inserted would be heterozygous for both the 'wild type' copy of the allele and the transposon interrupted copy. To test for the genomic integration involved in transposition, each of these eight clones was sporulated and then dissected for tetrads. Upon replica plating from the rich media to selection media for the *KanMX* transposon payload, I observed the expected 2:2 segregation expected of a genomically integrated transposon [Figure 4.2.2.]. This confirmed that the transposon system worked as intended in my hands within an intra-species diploid. In fact, one of the clones had actually had a transposition into an essential gene, evidenced by the 2:0 segregation observed upon rich media dissection.

While this tactic of sporulation and tetrad dissection to confirm genomic integration was possible in the *S. cerevisiae* diploid, the sterility of the *S. cerevisiae* x *S. uvarum* hybrid provided a barrier to this. As a workaround to this barrier, I adapted a transposon integration sanger sequencing method developed by the graduate student Wei Zhou in the lab of Stanley Fields in the department of Genome Sciences. The method is described within the Method section of this

chapter, but briefly genomic DNA was isolated from each of the eight transposon integration clones of the hybrid. This genomic DNA was then digested by two common six cutter restriction enzymes to fragment the genomes. The next step was a diluted intramolecular ligation, followed by purification and submission of these ligation populations to sanger sequencing with primers that read out of the left arm of the transposon thus capturing the transposon-genomic DNA junction at the site of integration. Each of these clones showed integration events of the transposon within both the *S. cerevisiae* and *S. uvarum* parental genomes, proving the efficacy of the method for my system of interest [Figure 4.2.2.]. With this proof in hand, I moved forward to generate the necessary homozygous knockout hybrid strains for three different differentially essential candidate genes.

Figure 4.2.2. Sanger sequencing quality traces from hybrid transposition colonies. Each trace is from a different colony, with the genomic locus of insertion labeled.



4.2.2 Selection process for three differentially essential gene candidates that are *S. cerevisiae* species specific: *MYO2*, *SEC24*, *TFC3*.

The work described here focuses on differences in gene essentiality between two *Saccharomyces* species, but previous work in the field has also shown that even within different strains or

isolates of *S .cerevisiae* there are observed differences in essential genes. One body of work compared the list of essential genes between the lab strain S288C and another isolate named Sigma1278b that have ~0.3% sequence divergence and found several dozen that were essential in either one isolate or the other. This work, and several other studies since, have shown consistently that there is some degree of variation in essential genes between isolates: a rather interesting and confounding result since one would not expect essential genes to change between isolates of the same species[142].

This work laid the foundation for my own thesis work on studying differential gene essentiality between *Saccharomyces* species, but also highlighted the importance of carefully selecting candidate genes that are *S. cerevisiae* species specific. If a candidate is chosen that is differentially essential even within *S. cerevisiae* isolates, results could be confounded by this isolate by isolate background variation in addition to species level differences between *S. cerevisiae* and *S. uvarum*. With this in mind, I aimed to comb through literature within the community to select candidates from what I describe as core essential genes: those genes that are deemed essential in every *S. cerevisiae* isolate data has been recovered for.

From the heroic efforts of groups such as Joeseph Schacherer and Jolanda van Leeuwen's, there exist many large datasets profiling lists of essential genes for several different lab and natural *S. cerevisiae* isolates. While there is some degree of variation, there are indeed a significant number of essential genes that can be labeled as core essential. I immediately eliminated any candidate *S. cerevisiae* specific essential genes that were conditionally essential between strains. One obstacle to collecting multiple different datasets however is that methods

for interrogating essential genes can vary.

While some studies unambiguously generate deletion strains en masse similar to the original yeast deletion collection, other groups would create computational predictive models that could assign gene essentiality by growth in different conditions or expression level changes. These approaches I labeled predicted essential, and given the large list of initial *S. cerevisiae* specific

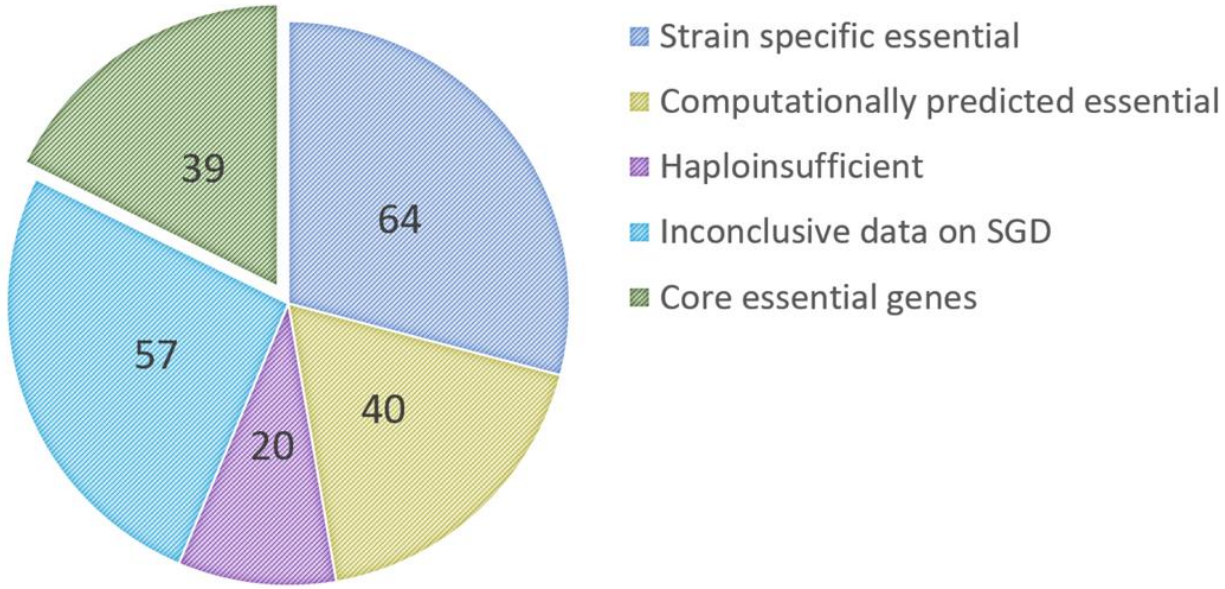


Figure 4.2.3. Pie chart showing screening process to generate final list of core *S. cerevisiae* essential genes that are differentially essential with *S. uvarum.*

essential genes originally predicted and experimentally verified by a previous graduate student in the lab, I eliminated all genes that were predicted essential only, without experimental data. In order to avoid any potential pitfalls or experimental difficulties, I then eliminated any remaining core essential genes that were haploinsufficient, or whose essentially was contradicted by publicly available datasets found on the community resource the *Saccharomyces* Genome Database (SGD) [Figure 4.2.3.][143].

These filters left me with a final list of thirty-nine core essential genes that were also predicted *S.*

cerevisiae specific essential genes in the graduate student's data. As a final stringent factor, I then only selected genes from this list of 22 genes that the previous student had experimentally verified the differential essentiality for between *S. cerevisiae* and *S. uvarum*. Unfortunately, given the smaller amount of data generated for *S. uvarum* studies compared to *S. cerevisiae*, these candidate genes have only been tested for essentiality in a single *S. uvarum* strain background. I was left with my final three candidate genes I had chosen from the remaining pool based on interest in function: *MYO2*, *SEC24*, and *TFC3*. *MYO2* encodes a Type V myosin motor involved in actin-based transport of cargos, required for the polarized delivery of secretory vesicles[9,10]. I am interested in this gene specifically because it has many different interacting partners in its multiple pathways, all of which could potentially show up as hits for interacting loci changes between *S. cerevisiae* and *S. uvarum*[11]. For the other two gene candidates, *SEC24* encodes a component of the Sec23p-Sec24p heterodimer of the COPII vesicle coat, required for cargo selection during vesicle formation in ER to Golgi transport[12].The final selected candidate gene was *TFC3*, which codes for a subunit of the RNA polymerase III transcription initiation factor complex[13]. Again, all of these candidate genes were selected from an experimentally validated list of core *S. cerevisiae* essential genes that were nonessential in *S. uvarum*, and whose biological function prompts ideas of potential hits in my genetic mapping approach.

4.2.3 Testing of conditional *S. cerevisiae* strains to generate haploid conditional shut-off strains for essential gene candidates.

Before I describe how I successfully engineered my *S. cerevisiae* x *S. uvarum* strains, I will first begin by detailing several other strain engineering methods I attempted which were ultimately

unsuccessful. The generation of haploid knockouts for the genes of interest in *S. cerevisiae* was quite a bit trickier and required leveraging the awesome wealth of resources available to the yeast community. This was because all three of the candidate genes were essential in *S. cerevisiae*, but not in *S. uvarum*, meaning that attempting to generate a knockout of these genes within a haploid *S. cerevisiae* strain would render the resulting knockout strain inviable. Getting around this inviability requires leveraging yeast resources such as the heterozygous knockout strains from the deletion collection generated from the efforts of many labs within the budding yeast scientific community. In addition to the deletion strain collection, partial collections of conditional shut off strains leveraging promoters responsive to drugs such as Tetracycline or doxycycline have also been generated.

Conditionally responsive strains provide an alternative method to achieving a homozygous knock out of a differentially essential gene within the *S. cerevisiae* x *S. uvarum* hybrid. Instead of strain engineering techniques and classical yeast methods such as tetrad dissection, conditional *S. cerevisiae* haploid strains can simply be mated to their corresponding *S. uvarum* gene deletion partner and then grown on the correct media for repression of the essential gene expression. In this way, conditional expression collections provide another method to work with haploids that can then functionally be treated as LOF alleles for the *S. cerevisiae* essential gene deletion. The specific collections which I tested were the yeast Tet-promoters Hughes Collection (yTHC), and a β-estradiol responsive collection from Scott McIsaac et. al. [Figure 4.2.4.]. The yTHC collection relies upon a tetracycline shut off system where the native promoter of a gene of interest has been replaced by a tandem repeat of sequences making up the TetO promoter from bacteria[144,145]. Because of promoter replacement, an essential gene will be expressed on rich

media and its expression shut-off when grown on media containing doxycycline—a more stable

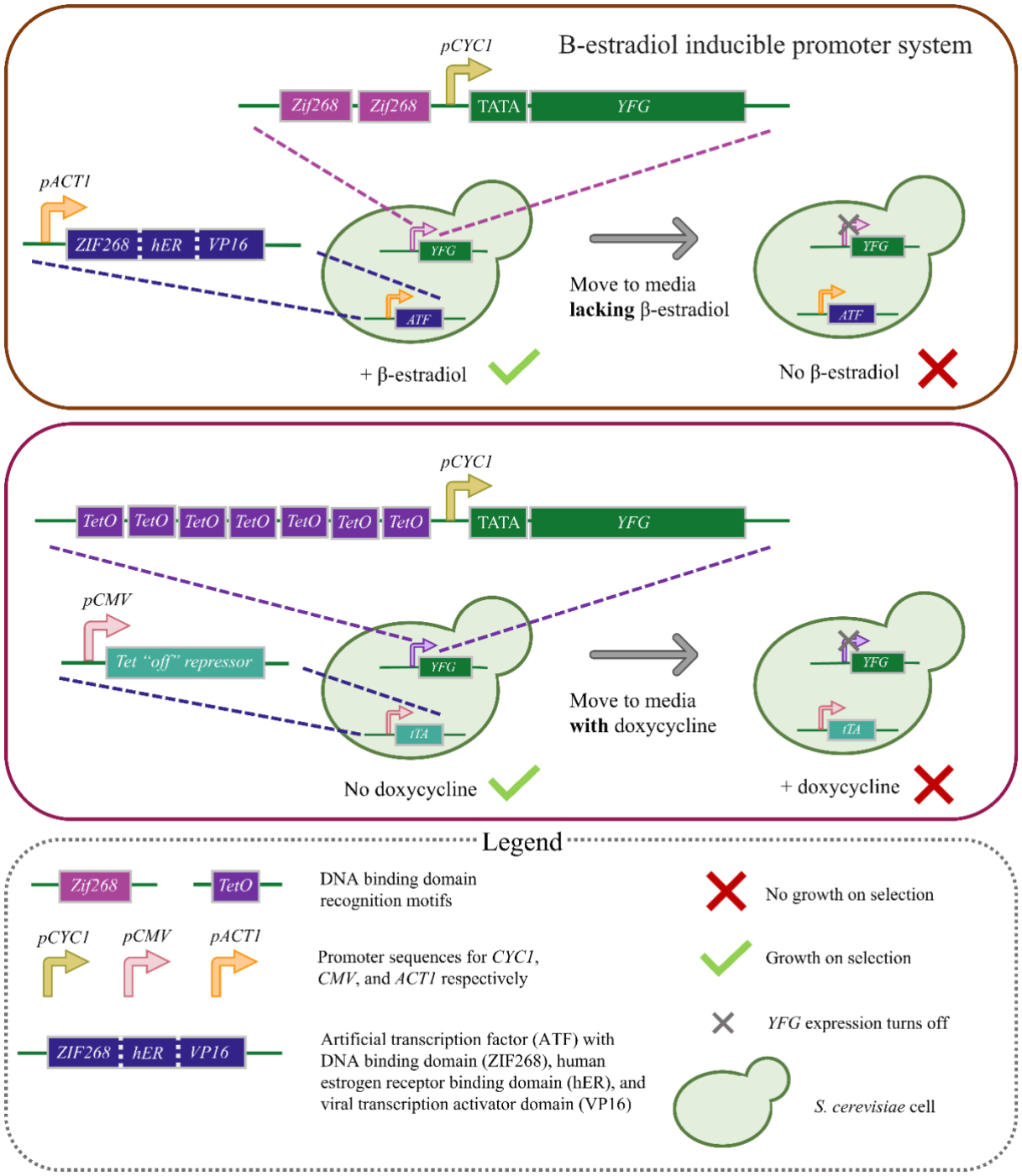


Figure 4.2.4. Schematic depicting both titratable promoter systems tested: β-estradiol and doxycycline. The first is an inducible system where β-estradiol must be present for a haploid essential conditional strain to grow. The latter is the reverse: adding doxycycline shuts off expression of the essential gene and leads to cell death.

version of tetracycline[146]. This conditional shut-off of expression results in mimicry of a LOF knockout phenotype, and thus any haploid yTHC essential strains should not be able to grow on media containing doxycycline. In my tests of efficacy for a few, select candidate genes specifically essential in only *S. cerevisiae*, titrating doxycycline concentration up to ten times the recommended dosage of 10 µg/mL still did not result in complete shut off for any of the essential gene candidates [Figure 4.2.5.].

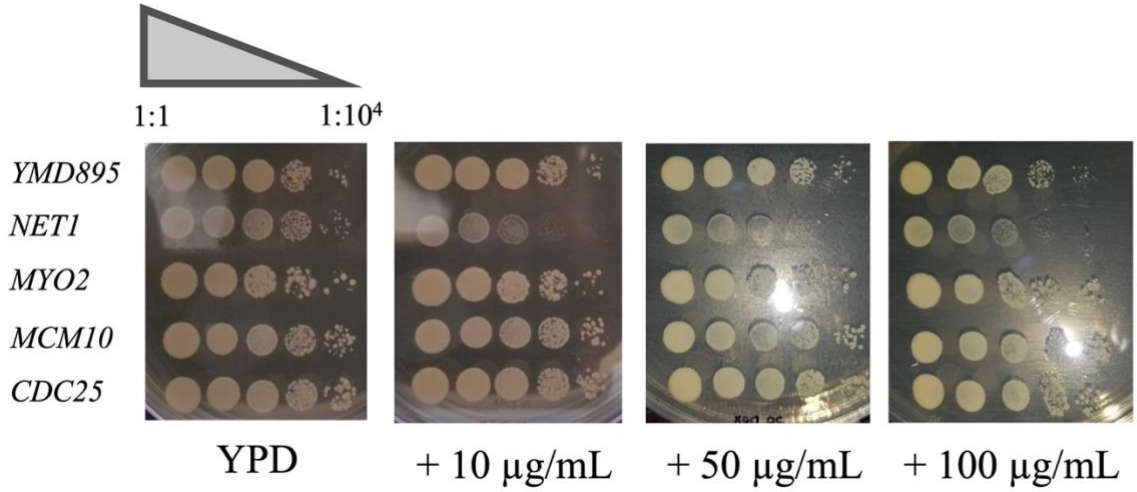


Figure 4.2.5. Spotting assays for select *S. cerevisiae* species specific essential gene strains from the yTHC collection on differing concentrations of doxycycline. Serial dilutions start from OD 1.0 of 1:1, 1:10, $1:10^2$, $1:10^3$, $1:10^4$.

For the β-estradiol responsive collection, the system works in the opposite direction[147]. Addition of the drug, in this case the human derived hormone β-estradiol, results in expression of a synthetic promoter through the recruitment of an exogenous, designed artificial transcription factor (ATF). This ATF contains zinc-finger DNA binding domains specific to its corresponding designed promoter which has replaced the endogenous promoter for the gene of interest. The transcription factor also consists of a human estrogen binding domain as well as a viral activation domain known as VP16. Without β-estradiol, the estrogen receptor domain of ATF interacts with

the Hsp90 complex and remains sequestered in the cytoplasm. Once β-estradiol is added to the media, Hsp90 interaction is displaced and a nuclear localization signal for the ATF is revealed and it can then induce expression of the gene of interest after translocation into the nucleus. This system together means that for a haploid *S. cerevisiae* strain with a promoter replacement at an essential gene, β-estradiol is necessary in the media for viability and growth.

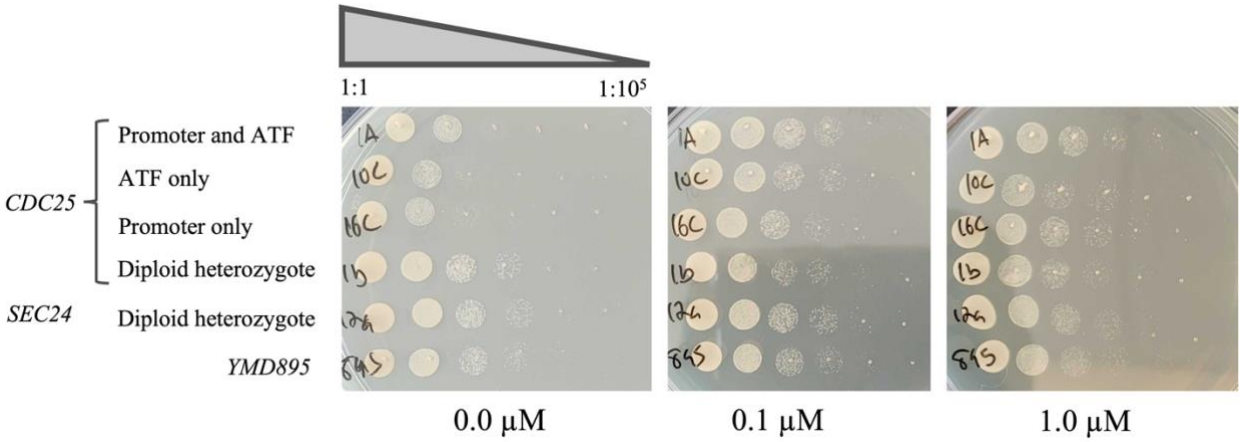


Figure 4.2.6. Spotting assay for β-estradiol conditional essential gene strains for *S. cerevisiae* species specific candidate genes on different concentrations of β-estradiol. Serial dilutions start from OD 1.0 of 1:1, 1:10, $1:10^2$, $1:10^3$, $1:10^4$, $1:10^5$.

In this design, the β-estradiol conditional *S. cerevisiae* strain is mated to its *S. uvarum* partner with its own differentially essential gene ortholog knocked out. After the hybrid is formed, it can be grown on media lacking any β-estradiol, leading to shut-off of the essential gene's expression, again achieving the same phenotypic effect of a LOF mutation in the essential gene.

Unfortunately, as with the yTHC collection, titrating β-estradiol in growth media for several candidate haploid strains with promoter replacement at essential genes led to no loss of viability without β-estradiol [Figure 4.2.6.]. While the results of both of these titratable promoter strain collections were unfortunate, the noisiness and imperfections of inducible promoter systems within the context of a cell have been observed previously[148–150]. From experience in our lab, as well as discussions in published work, generally it seems that these titratable strains work or do

not work on a gene by gene basis. While for some genes residual expression in the shut off condition is enough to maintain cell viability, for others any leaky expression is still not enough to rescue strain inviability from losing normal expression of an essential gene.

4.2.4 Generation of hybrid strains with homozygous knockouts of differentially essential candidate genes leveraging strain engineering

The knockout *S. uvarum* haploids were generated using homologous recombination as described previously. Briefly, a 'wild type' diploid *S. uvarum* strain was transformed with a *HygMX* cassette flanked by homology sequence for the *MYO2*, *SEC24*, and *TFC3* loci respectively. These heterozygous knockouts were then sporulated and haploid *S. uvarum* strains bearing *HygMX* deletion cassettes at each of the candidate gene loci were recovered. It was also important to ensure all strains, both *S. cerevisiae* and *S. uvarum*, also bore knockout alleles for *URA3* for the counter selection of the *URA3* gene on the transposon vector.

There was also another consideration to generating these hybrids in addition to the differential essential gene knockouts: mitochondrial inheritance. After hybrid formation, mitochondrial-nuclear genome incompatibilities and effects depending on which parental mitochondrial genome is present in the hybrid have been observed before[151–153]. Research has shown that in the populations derived from a single *Saccharomyces* hybrid mating event, each cell does not contain a mixture of both parental mitochondrial genomes, but instead experience uniparental inheritance[101,154,155]. Thus, for a *S. cerevisiae* x *S. uvarum* hybrid population, each cell either contains only *S. cerevisiae* mitochondrial genomes or only *S. uvarum* ones. There is recent evidence that the uniparental inheritance of mitochondria can even affect transposition rate

efficiencies in *Saccharomyces* hybrids

To avoid complications involved in heterogenous mitochondrial inheritance within a hybrid population, all *S. uvarum* strains were cured of their mitochondrial DNA (mtDNA). Curing was performed by leveraging an ethidium bromide curing protocol, which specifically inhibits the polymerases that function in the mitochondria and thus leads to mtDNA loss from the cell. The loss of mtDNA following growth in ethidium bromide was confirmed by their inability to grow on the non-fermentable carbon source glycerol [Figure 4.2.7.]. This specific curing strategy ensures all mtDNA inherited within the *S. cerevisiae* x *S. uvarum* hybrids in this body of work only have *S. cerevisiae* mitochondrial genomes.

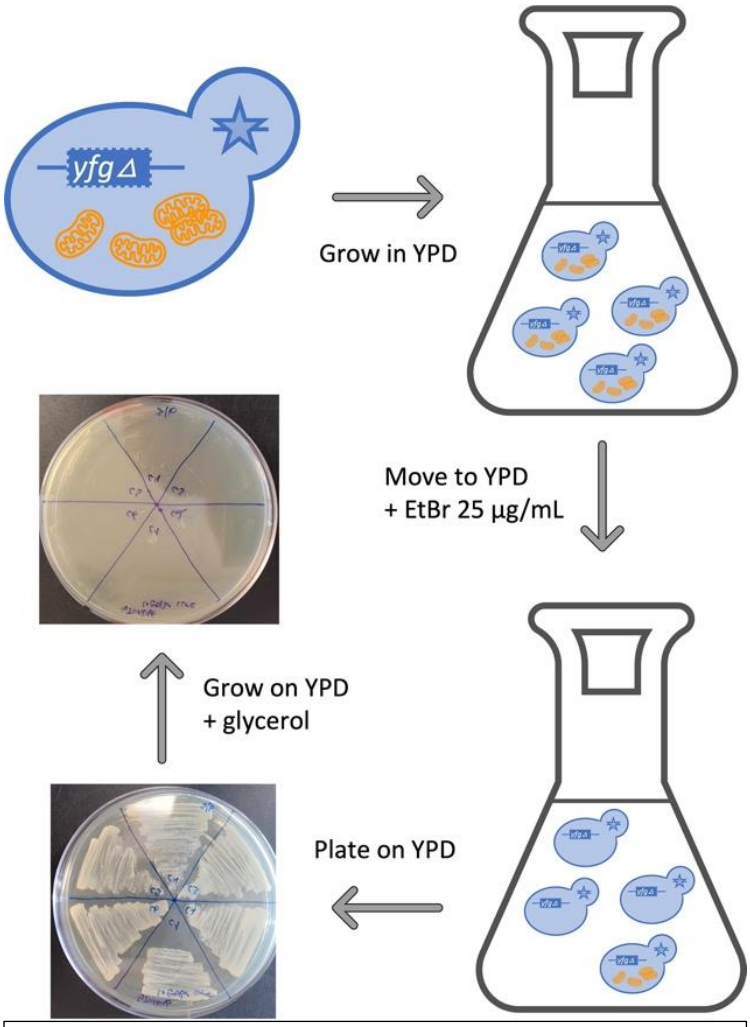


Figure 4.2.7. Experimental diagram of ethidium bromide curing of mtDNA from all *S. uvarum* differentially essential gene knockout strains.

To overcome the obstacles present when working with haploid knock out *S. cerevisiae* strains of essential genes discussed previously, I turned to the full gene deletion collection as a resource to

engineer the hybrid strains necessary for RH-seq transposon mutagenesis. These strains were originally subjected to drug marker replacement transformation first to avoid issues with the *KanMX* cassette—which confers resistance to kanamycin and geneticin—being used as the transposon payload and the drug marker for the essential gene knockout. Further engineering was performed by leveraging another community resource known as the MoBY collection. The MOBY collection has two versions, a low and a high plasmid copy number, each consisting of a pool of barcoded 'wild type' versions of all *S. cerevisiae* genes on an exogenous plasmid, and this resource was also of great use here[156]. With it, I was able to transform the marker swapped, *NatMX* cassette—conferring nourseothricin, or clonNAT drug resistance—marked heterozygous knockout *S. cerevisiae* strains with the corresponding 'wild type' allele plasmid. These plasmid containing heterozygous knockouts of essential gene strains were then sporulated and dissected on media selective for plasmid maintenance. After replica plating to media containing cloNAT to select for the essential gene knockout, haploid *S. cerevisiae* strains were then recovered that

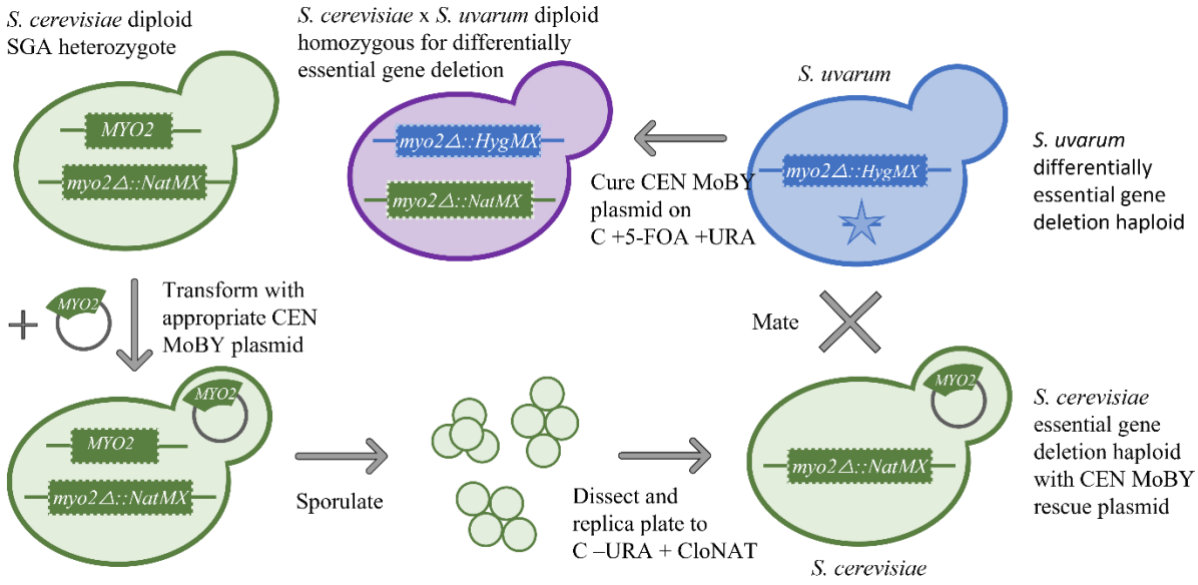


Figure 4.2.8. Strain engineering approach for *S. cerevisiae* essential gene deletion haploid strains. Protocol was done for each candidate *S. cerevisiae* species specific essential gene.

contained a knockout allele of an essential gene that was rescued by the presence of the corresponding MOBY collection plasmid.

The final transposon mutagenesis strains were then constructed by mating the haploid *S. cerevisiae* knockout with the rescue plasmid to the corresponding *S. uvarum* strain containing the same differentially essential gene knockout [Figure 4.2.8.]. After mating, the rescue plasmid from the *S. cerevisiae* strain was then cured out by utilizing the *URA3* counter selectable marker. It was not known previously that this hybrid would be viable: a homozygous knockout for both parental alleles of the differentially essential gene. Since all three candidate genes were viable as hybrids with a homozygous knockout for the candidate gene, this suggests that the rescue factors are dominant to the *S. cerevisiae* non rescue alleles. The strain engineering was then complete for *MYO2*, *SEC24*, and the *TFC3* candidate genes as well as a 'wild type' hybrid of the same genetic backgrounds but with no differentially essential genes knocked out.

4.2.5 Construction and sequencing of transposon mutagenesis libraries in *myo2Δ*, *sec24Δ*, *tfc3Δ*, and 'wild type' *S. uvarum* x *S. cerevisiae* hybrids.

Transposon mutagenesis in each of the four hybrid strains was performed as previously described[141]. Briefly, each strain was transformed with the transposon plasmid pJR487 en masse in large culture volumes in order to recover many independent transposition events along the genome. This means that ideally, enough unique transposition events occur in the pool of cells to ensure that each genomic locus has undergone multiple independent transposition events. The recovery of enough cells to achieve this transposon insertion rate coverage empowers the method for detection of statistically significant differences in recovered transposition events between

'wild type' and experimental strains within the two parental genomes [Figure 4.2.9.]. The construction of all four libraries was completed as biological replicates to control for any experimental noise between library construction days.

These libraries were then each individually prepared for transposon enrichment sequencing as previously described. Briefly, genomic DNA from a transposon mutagenesis library was sonicated with the Covaris machine to roughly 300 bp fragments. This sheared DNA was then

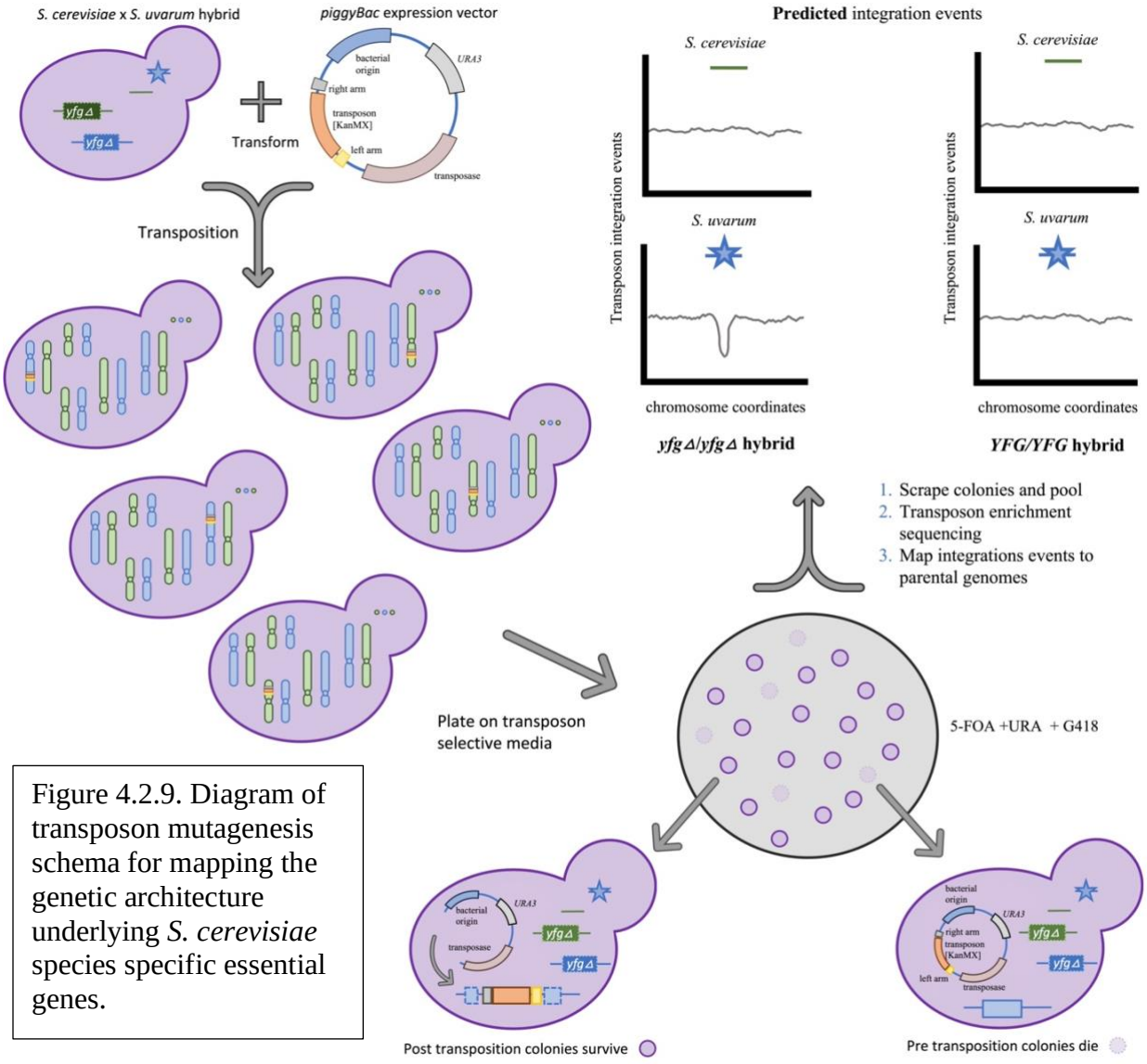


Figure 4.2.9. Diagram of transposon mutagenesis schema for mapping the genetic architecture underlying *S. cerevisiae* species specific essential genes.

subjected to end repair, A-tailing, and adaptor ligation followed by a transposon specific PCR as described in earlier studies and in the RH-seq paper from the Brem group. This PCR step is critical in order to obtain and sequence only the transposon-genomic DNA junctions to then map insertion sites, which was done computationally as described below. In addition, this PCR step also adds the P5 and P7 sequences critical for cluster generation on an Illumina sequencer [Figure 4.2.10.]. Each library was sequenced on an Illumina NextSeq 550 with paired end 150 bp reads using TruSeq sequencing primers.

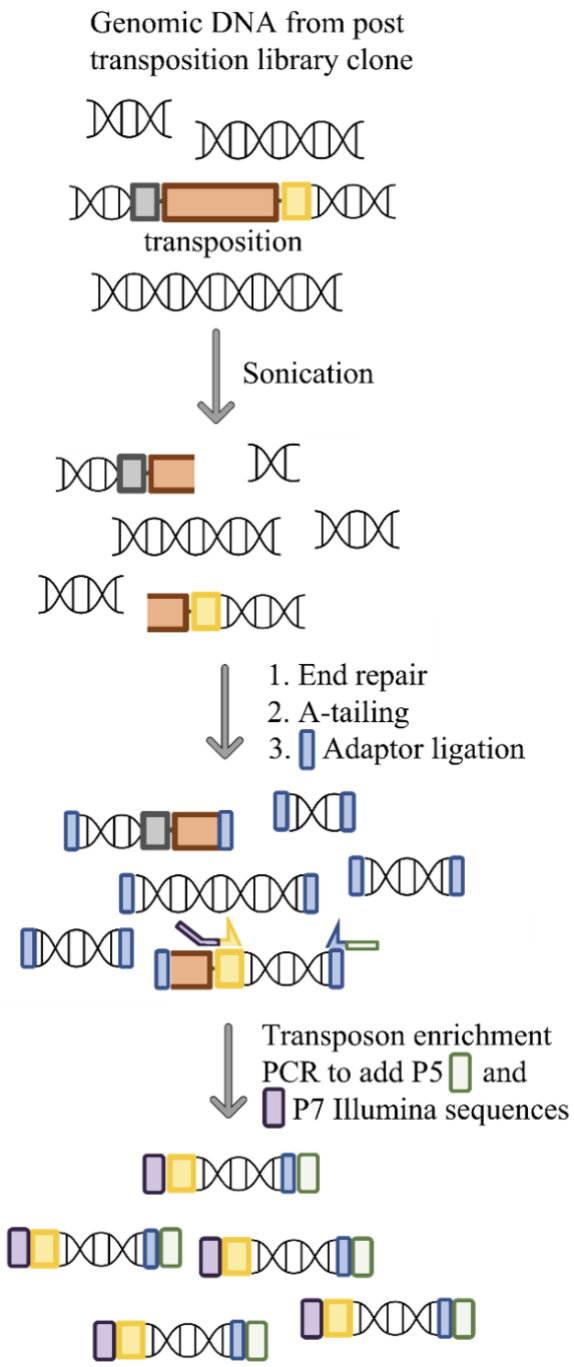


Figure 4.2.10. Experimental outline of transposon enrichment next generation sequencing library preparation.

4.2.6 Computational pipeline to call differences in transposon insertion rates between 'wild type' and differentially essential gene knockout hybrids.

Transposon enrichment sequencing required a length time period of troubleshooting due to the challenging nature of performing each sequencing preparation step independently. Recent evidence has suggested that this

protocol has begun to work. Unfortunately, there still remain some issues with consistency of fragment length after the transposon enrichment PCR, further leading to poor quality of Nextera sequencing reads. While this protocol is still being subjected to final troubleshooting, a computational pipeline has been identified to facilitate swift analysis of recovered transposition events between experimental and control hybrid strains. The planned analysis leverages an existing computational pipeline within the Dunham that has been applied to copy number variation. The pipeline requires a previous alignment of sequencing reads to the appropriate genome, which results in a file type known as .bam. In the cast of the transposition enrichment sequencing, the reads would first need to be trimmed of the transposon sequence adjacent to the transposon-genome junction. After trimming, these reads should align to either the *S. cerevisiae* or *S. uvarum* genome contained in a .gff genome file type used for alignment. After this alignment of the transposon adjacent reads to the hybrid genome, the copy number variation pipeline works by relating read number, also referred to as read depth, to genomic copy number.

In the original application of this pipeline within the lab, a strain that has had a duplication event for gene A will have approximately twice the number of reads mapping to that genomic locus when compared with a strain that has not undergone a duplication event and only has a single copy of gene A. In the case of transposon events, for a control versus experimental hybrid strain, we expect a greater number of reads to be observed at a rescue locus only in the control strain. This is because more insertion events in the population for a gene will be reflected as a greater number of reads when compared to the experimental strain. Specifically, when compared to the experimental strain for the rescue loci expected to be mapped within the *S. uvarum* parental genome. Thus, mapping of rescue loci can be performed by leveraging this copy number pipeline

which will convert read depth of transposition events to number of transposon insertions within the population.

4.3 Discussion

It is unfortunate that final sequencing data has not been obtained yet for any of the transposon mutagenesis experimental or control libraries. However, it is possible to think forward and speculate on what the results could look like. One point that has been discussed previously but bears mentioning again is the genetic complexity underlying differential gene essentiality. After sequencing and analysis by the copy number pipeline, it is likely that there will be more than one rescue locus responsible for differential gene essentiality. The challenge with drawing conclusions about the contributions of each locus to the suppression of essentiality in *S. uvarum* is the degree to which each locus contributes. If the underlying cause of differential essentiality is many contributing loci of small effect, the mapping of this genetic architecture becomes more challenging. Depending on the size of the effect, it can be difficult to make decisions for cut offs between experimental noise and actual effect for the mapping of these loci.

However, part of the reason we chose several differentially essential candidate genes was the hope that among the candidates there would be varying complexity of the underlying architecture. While one candidate could have very complex architecture underlying the *S. cerevisiae* exclusive essential gene, another could have more simple architecture that is more amenable to genetic mapping. Another way in which we have worked to mitigate the possibility of these mapping challenges is the choice of method. RH-seq transposon mutagenesis has been shown previously to be capable of mapping more complex genetic traits. By leveraging yeast

population sizes to generate large transposon mutagenesis libraries and sequencing depth and analysis the Brem group was able to identify many underlying loci for their phenotype of interest—thermotolerance. In addition, some of these loci had the small effect sizes that have been discussed previously here. Even with the challenges of analysis to identify the rescue loci of varying complexity, we hypothesize that experimental design and choice of tool will help overcome these challenges and enable the mapping of rescue loci underlying differential gene essentiality. By mapping the genetic architecture underlying differential gene essentiality between *S. cerevisiae* and *S. uvarum*, our work will provide more insight into the effect of genetic background on the phenotype of gene essentiality, and more broadly increase the field's understanding of how local genetic interactions evolve over time.

4.4 Materials and methods

Strains

All strains used came from a parental genetic background of S288C for *S. cerevisiae* and CBS7001 for *S. uvarum*. Haploid strains for both species with each gene of interest knocked out (KO) were mated to create the final hybrid strain for transposon mutagenesis. The construction of the haploid *S. cerevisiae* KO strains was as follows: the heterozygous knockout diploid from the Yeast Deletion Collection was transformed with its corresponding CEN MoBY collection plasmid using a 1-step transformation protocol. Briefly, the strain was grown up in 5 mL of YPD and then 1 mL of this culture was spun down. After a wash step with water, 1 μL of miniprepped plasmid was added to the pellet followed by 75 μL of 1-step transformation buffer. Cells were then incubated at 42°C for 1 hour and then plated on C -Ura plates. This heterozygous diploid strain transformed with the rescue plasmid was then sporulated in 5 mLs of 1% potassium

acetate, 0.05% dextrose, and 0.1% yeast extract for 4-7 days. Sporulated cultures were then dissected directly onto C -Ura plates to maintain the CEN MoBY plasmid. Dissected tetrads were then replica plated to D +CloNAT to select colonies that were Ura+ and NatMX+. These haploid *S. cerevisiae* strains with the KO allele and rescue plasmid were then mated to the appropriated *S. uvarum* haploid KO strain. These *S. uvarum* haploids were generated as follows: primers designed with 40 bp of homology to the gene of interest were used to amplify the *HygMX* cassette from pRS40H. The parent *S. uvarum* strain YTW8 was then transformed with purified PCR product as described previously. After sporulation, haploids with the KO allele were verified by PCR and then mated to a *ura3Δ S. uvarum* haploid. Following another round of sporulation and dissection, haploid *S. uvarum* were isolated with the desired gene KO and Ura- phenotype. To avoid incompatible mitochondrial inheritance, all *S. uvarum* strains were then subjected to mitochondrial DNA curing. Cells were grown overnight in 5 mL YPD. These cells were then transferred to a culture of 25 mL YPD at an amount for a final OD of 0.2 with 25 μg/mL of EtBr added. After overnight growth serial dilutions of the culture were plated on YPD. After several days of growth, colonies were replica plated to YEPG (glycerol containing medium) and screened for inability to grow on glycerol containing media. Hybrids were created by patch mating the corresponding *S. cerevisiae* and *S. uvarum* KO strains and selecting for diploids via drug and plasmid selections. After several days of growth, these mated strains were then streaked out onto 5 FOA plates to select for colonies which had lost the CEN MoBY rescue plasmid from the *S. cerevisiae* strain. The final strains were then used for transposon mutagenesis.

Growth

Cultures of pJR487 prepared for Zymo Research bacterial Gigaprep kits were grown at 37°C in LB media with 200 µg/mL of Kanamycin. *S. cerevisiae* strains were all grown at 30°C, while all *S. uvarum* strains and *S. cerevisiae* x *S. uvarum* hybrids were grown at 25°C.

Transposon mutagenesis

Before transposon transformation, pJR487 was gigaprepped using a column kit (Zymo Research) to generate about 11 mg of plasmid. For the strain to be transformed, it was streaked out from glycerol onto YPD and grown at 25°C for two days. The day before transformation, a single colony was inoculated into 100 mL YPD and shaken at 25°C, 200rpm for ~24 hours. Cells from this pre-culture were then transferred to each of four 1 L flasks at the volumes required to attain an optical density at 600 nm (OD600) of 0.2 in 500 mL YPD each. These cultures were grown for 6 hours at 25°C with shaking at 200rpm. After growth, two of these 500 mL cultures were combined into 1 L of culture, and the last two 500 mL cultures were combined into a separate 1 L. Each of the two 1 L cultures were transformed (for a total of two transformations) as follows: cells were split into twenty 50-mL conical tubes and centrifuged at 4000 RPM for 5 minutes. The cells were then washed with 25 mL of water per conical. This wash was repeated with 0.1 M lithium acetate (LiOAc) mixed with 1X Tris-EDTA buffer (10 mM Tris-HCl and 1.0 mM EDTA). After the wash, the following mixture was added to each conical: 1) 0.269 mg of pJR487 mixed 5:1 by volume with salmon sperm DNA, 2) 3 mL of 39.52% 3350 MW polyethylene glycol, 0.12M LiOAc and 1.2X Tris-EDTA buffer (12 mM Tris-HCl and 1.2 mM EDTA). Cells with transformation mix were incubated for 10 minutes at room temperature and then heat-shocked in a water bath at 37°C for 26 minutes. Cells from all 20 tubes were then combined and transferred to each of three 1 L flasks at the volumes required to attain an OD600

of ~0.35-4 in 500 mL of YPD. Transformed cells were recovered by shaking at 28°C and 200 rpm for 2 hours, after which G418 (Geneticin, Gibco) was added to each flask at a concentration of 300 μg/mL to select for those cells which had taken up the plasmid. Post-transformation cultures were grown with 200 rpm shaking at 25°C for two days until each reached an OD600 of ~14.5. All 6 recovered cultures were then combined into one, then cells from this combined culture were transferred to each of two 1 L flasks containing YPD + G418 (300 ug/mL) at the volumes required to attain an OD600 of 0.2 in 500 mL each. Cultured each flask at 25°C and 200 rpm shaking overnight until reaching an OD600 of 6.0. The two cultures were then combined into one culture, and a volume of this master culture was spun down and resuspended in water to attain a cell density of 1.85 OD600 units/mL in 12 mLs. This resuspension was then plated on complete synthetic media with 5-fluoroorotic acid (5-FOA) and G418: 340 μL was plated on each of a total of 36 large, 15 cm diameter plates. After incubation for two days at 25°C to enable colony growth, colonies were scraped off of all 36 plates and combined into water at the volume required to attain 40 OD600 units per 900 μL. This suspension was then aliquoted into 12 total glycerol stocks of 1.5 mLs of library and 0.5 mLs of 50% glycerol. This is the final transposon mutant hemizygote library.

Transposon sequencing library construction

To generate Illumina transposon enrichment sequencing libraries, 300 μL of the final concentrated library stock of OD ~100 was subjected to a genomic DNA preparation protocol as described previously[157]. In short, the cells were washed with water, and then transferred to a lidlock or otherwise securely closed microcentrifuge tube. To these cells were added 300 mg of acid-washed glass beads, 200 μL of lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10

mM Tris-HCl (pH = 8), 1 mM EDTA), and 200 μL 25:24:1 phenol:chloroform:isoamyl alcohol. After 8 minutes of vortexing, samples are spun down and resuspended in 200 μL TE buffer and the top aqueous layer transferred to a new tube containing 1 mL 100% ethanol and spun down again. The pellet is then resuspended in 400 μL TE with 30 μg of RNase A and incubated at 37°C for 30 minutes. After incubation, 10 μL of ammonium acetate and 1 mL of 100% ethanol were added to finish the DNA precipitation. Samples were then resuspended in 50 μL of TE and quantified using a Qubit fluorimeter.

One microgram of cleaned and quantified gDNA was placed in a Covaris sample tube and sheared on the machine for an average final length of 300 bp fragments. After AMPure clean ups described previously, each sample was then subjected to end repair, A tailing, and adaptor ligation[158]. A transposon enrichment PCR amplification was then carried out with the transposon specific primer TW195 and the appropriate indexed reverse primer. These post-PCR libraries were then AMPure purified and quantified on a Qubit using the High Sensitivity (HF) protocol. Libraries were then also run on 6% TBE vertical gels at 150 V for 40 minutes, then stained for 5-10 minutes with SYBR gold before being imaged on an UV transilluminator to check for band smear size and library quality. Good quality libraries were then pooled at 2 nM each and run on an Illumina NextSeq 500/550 as 150 bp paired end reads. Due to the repeated sequence at the beginning of these libraries and their low complexity, these transposon enrichment sequencing libraries could not take up more than 15% of the total reads for a Nextera sequencing kit.

Chapter Five: Alternative methods for genetic mapping in sterile hybrids and future directions

5.1.1 A brief discussion on considerations for transposon mutagenesis limitations.

While an extremely powerful method that has been shown to be up to the task of many different styles of biological questions, transposons are not a perfect tool. When using transposon mutagenesis, it becomes imperative to understand the drawbacks and biases of the specific transposon of choice. Certain transposons like piggyBac are preferrable due to their scarless insertion, and the very general preference of "TTAA" sequence for a site of insertion[83,84,159]p. Other transposons such as Ac/Ds suffer from drawbacks such as genomic scarring at the site of insertion, while others still have more specific insertional sequence bias. This is especially important because of how a transposon mutagenesis approach works: the number of potential insertion sites functions as a limiting factor for the percentage of the genome whose function can even be interrogated by transposon insertion. If a gene does not contain the sequence, there is no insertion and thus no information can be gathered about the effect of transposition LOF on such a gene.

Another important consideration is that transposition itself is a low percentage event. Other transposon systems like SATAY that function in a way that allows for specific measure of transposition efficiency has shown this number varies but is generally around 1%. While it is possible to overcome such a limitation by sheer number of cells, as done in this work, it is important to keep in mind that this system becomes less powerful or useful as a method in other model organisms where it is not so easy to grow that magnitude of cells. And even where it may be possible for a lab to have enough tissue culture hoods to expand a human cell line to such a

degree, space and financial limitations mean this work can only be carried out only in the most highly funded areas of science. The piggyBac method was a good choice as a tool for our question here, but future scientists should exhibit equal care with choosing the right tools for the problem at hand. You can never answer a biological question you do not have the ability to interrogate from lack of proper tools: like the gene with no transposon preference insertion site.

5.1.2 Alternative approaches to mapping the genetic architecture of differential gene essentiality in *S. cerevisiae* vs *S. uvarum*.

Instead of transposon mutagenesis, several alternative methods exist for mapping the genetic architecture of differential gene essentiality. The selection of focusing on *S. cerevisiae* specific essential genes was a part of this calculus, because of the existence of many different community strain collections and other genetic resources. However, there are a few approaches that are better undertaken for *S. uvarum* species specific genes. One such approach leverages a *S. cerevisiae* resource mentioned previously in this work, the Cen MoBY collection. Each individual *S. cerevisiae* gene is expressed on a plasmid with a yeast CEN sequence that has each gene tied to a barcode sequence for ease of pooled assays. Armed with this collection, one could assay for interacting loci only in the *S. cerevisiae* that rescue essentiality of the differentially essential gene.

To do such an assay, one would individually generate several different heterozygous knockout *S. uvarum* strains for the gene of interest. Each separate candidate would be transformed en masse with the entire CEN MoBY collection of plasmids, such that each cell in a population has one CEN MoBY plasmid with the ‘wild type’ copy of a *S. cerevisiae* gene on it. To find interacting

factors specific to *S. cerevisiae* that rescues the gene essentiality in that species, these populations of *S. uvarum* heterozygous knock out strains carrying the CEN MoBY library would then be sporulated. After sporulation, cells would be subjected to stringent chemical and mechanical treatments in a random spore analysis protocol to select specifically for haploid spores on media that selects for the presence of the Cen MoBY plasmid as well as the *S. uvarum* deletion allele of the differentially essential gene. Surviving colonies from these mass platings would be scraped, pooled, and the barcodes sequenced using Illumina methods. Only those *S. uvarum* progeny carrying the 'wild type' *S. cerevisiae* genes belonging to the interacting network that rescues the essentiality in *S. cerevisiae* would be able to survive.

The identity of the barcodes present in the CEN MoBY plasmids of surviving progeny inform the identity of these interacting loci, upon which follow up studies could be performed. While this solution to genetic mapping in a sterile hybrid leverages powerful existing resources and using an elegant barcoding strategy, its major drawback lies in our lack of prior knowledge for the complexity of the genetic architecture for these interacting factors. Potentially this method could only work well for differentially essential genes where only a few interacting factors are involved, but struggle to be informative when the architecture becomes more complex. This was the primary reason I avoided this method, but I maintain such an approach could work in theory and would be worthwhile to attempt.

Another approach to mapping the underlying genetic architecture of differential gene essentiality leveraging another community resource would be working with chromosome replacement strains. Originally developed by H. Lee et al. the strains in question are *S. cerevisiae* haploid

strains containing single, or combinatorial replacement of their endogenous chromosomes with the corresponding *S. uvarum* chromosome[101]. The strains were created by the mating of the original haploid *S. cerevisiae* and *S. uvarum* parents, followed by sporulation and progeny recovery. The rare, few haploid progeny recovered were screened for the identity of each chromosome and those containing entire *S. uvarum* chromosomes were then backcrossed to the original *S. cerevisiae* parent on average five to eight times to finally obtain a haploid strain with entirely *S. cerevisiae* chromosomes except for individual replacements.

In the context of mapping the genetic architecture of differential gene essentiality, these strains could be used to specifically perform mapping for candidate genes essential only in *S. cerevisiae*. That is because of the experimental set up for how this mapping would be performed with the chromosome replacement strains. Leveraging the tool of rescue by mating as well, the method would work as described. First, several separate *S. cerevisiae* strains heterozygous for deletion of a species specific essential gene would be sporulated. This sporulation culture would then be subjected to RSA to destroy any remaining diploid cells. The recovered spores would be split into a number of aliquots, and each aliquot would be mated one by one to each of the different *S. cerevisiae* chromosome replacement strains. After mating, the surviving progeny would be plated on media selecting for the *S. cerevisiae* deletion allele for the species specific essential gene, and also for a growth media selection locus that can only be inherited from the *S. cerevisiae* chromosome replacement strains.

In this way, the only surviving progeny recovered would be those where the *S. cerevisiae* chromosome replacement strain mating partner contained the chromosome or chromosomes on

which the interacting loci present rescue the gene essentiality specifically in *S. uvarum*. While this method utilizes several clever yeast genetic methods, it unfortunately has several drawbacks that led me to set it aside as a method to perform the genetic mapping for these interacting loci underlying differential gene essentiality. Firstly, this approach would result in very coarse mapping initially which would then require many follow up experiments. Identifying the chromosomes or chromosomes that carry the interacting loci is a step towards mapping, but even with their relatively smaller size compared to human genomes, *Saccharomyces* chromosomes are still quite a large sequence space to have to search in finer detail through. In addition, not all of the single chromosome replacement strains were even able to be generated. Due to several reciprocal translocations in the *S. uvarum* genome compared to *S. cerevisiae*, these chromosomes were replaced as a pair. And finally, again the issue of how complex these interaction loci are arises. If the differential gene essentiality phenotype is comprised of a complex underlying genetic architecture spread across the genome, this method will lack the power to map these many loci of small to medium effect.

5.1.3 Building upon the work of generations of geneticists with an eye to the future.

One of the more enjoyable aspects of this thesis work has been traveling in many different regions of uncharted genetic territory. While the foundations of biology, genetics, and evolution were built upon thoroughly studied model organisms, in recent years it has become apparent that the modern descendants of these fields have begun to diverge from this fixed mindset. Looking back through scientific history, it made sense for the comparably few scientists of the day to focus their work on a few specific species. Primary in mind were factors such as generation time, ease of growth and maintenance, genome size, and eventually more modern ideas such as genetic

tractability and the ability to preserve strains. For decades work in classical modern organisms such as *E. coli*, *S. cerevisiae*, *C. elegans*, and *D. melanogaster* dominated the field of genetics[160]. Many classical studies were carried out in these organisms that revealed foundations in our understanding of fundamental biological processes such as genome replication, cell cycle control, cell lineage tracing, and even core features of multicellular organism development.

In present time, these restrictions of organisms in which studies are carried out have slowly been easing through a variety of factors. Key among them include the increasing accessibility of science to more than just a select, wealthy few. As the generation sizes of geneticists grew, the interplay of fields such as synthetic biology would also accelerate the development genetic tools. While previously the genetic tractability of a model organism was very important, hence *S. cerevisiae*'s importance for its high rate of homologous recombination for genome engineer, the development of widely applicable tools such as CRISPR/Cas9 has removed this obstacle[161,162]. With each scientific conference, more and more novel work is being presented on improving the function of genome editing in novel model organisms. Importantly, this increase in tools is being more than matched by a continuing escalation in the number of new assembled genomes being released to the scientific community[163–168].

The Human Genome Project to sequence an entire human genome may have been a paradigm shift of ground breaking proportions, but the scientists involved in this work must be truly amazed by the sequencing methods presently available to this newer generation of scientists[169]. What once took over a decade can now be done on a next generation sequencing machine within a few days, from the original isolation of genomic DNA. This is all without the need for

laborious cloning methods involving Bacterial Artificial Chromosomes (BACs). Even from the original invention of next generation sequencing there has been rapid growth: where once available methods could only sequence small fragments of the genome, newer methods such as PacBio make it possible to answer questions about copy number variation and tricky to sequence repetitive regions via the use of a long read sequencing method[170,171]. The fast paced growth of computational methods to parse and filter raw sequencing reads and generate a novel, assembled genome has also been stunning[172–176]. And while the development of the tools and technology has been an amazing result in and of itself, the possibilities all of these new tools present are truly astounding.

Freed from the idea of working within only a few, select model organisms, the potential genetic questions that can be asked are truly manifold. Truly, a study between 20 mya diverged *Saccharomyces* species as described here may one day be considered a relic of a past time. In the same way that the scientific period wherein the HGP was completed was full of fast paced discovery, the same feeling can be imagined to be captured now. With the increasing genetic accessibility of unique and interesting organisms like naked mole rats as a research organism, biological questions about important processes such as aging and the fundamental processes of cancer progression can be asked and answered[177,178]. The scope of what can now be asked is staggering, armed with all these tools and resources now available.

By detaching from other long held traditions similar to model organism like ‘wild type’, I believe the fields of biology and genetics can truly continue to undergo a rapid expansion in scope and ideas. Questions in human health and disease, the molecular underpinnings of evolution over

time for genetic networks, and even precision medicine targeting a drug treatment to a patient's specific ability to process such a drug, all of these questions are being considered in the current field. It truly feels like the next few decades will see a growth of knowledge and skills; that the scientific field is on the cusp of beginning to truly understand the biology of life around us. How genes working in tandem underlie complex phenotypes, how a single point mutation in a gene can affect human health and disease, all the way to a more complete understanding of the biology of the genome than we have ever had. It is an exciting time indeed to be a geneticist, I cannot wait to see what the next few decades have in store for this field built on tradition and the knowledge of everyone who has ever come before.

Appendix A

Strains

Strain name	Mating type	Genotype	Species
YTW1	Matα	Hodel::HisG Lys2 His3del1	*S. uvarum*
YTW2	Mata	Hodel::HisG LYS2 His3del1	*S. uvarum*
YTW3	Mata/Matα	can1D::STE2-Sp_his5/CAN1 lyp1D/LYP1 his3-1/his3delta1 leu2D0/leu2delta0 ura3D0/ura3delta0 met5D0 LYS2+/LYS2 MET15/met15delta0	*S. cerevisiae*
YMD895	Matα	can1D::STE2-Sp_his5 lyp1D his3-1 leu2D0 ura3D0 met5D0 LYS2+	*S. cerevisiae*
YTW125	Mata/Matα	Can1D::STE2-Sp_his5/CAN1 lyp1D/LYP1 his3-1/his3-delta1::CRE leu2D0/LEU2 ura3D0/ura3-delta0::CRE met5D0/MET5 LYS2+	*S. uvarum* x *S. cerevisiae*
YTW164	Mata/Matα	can1D::STE2-Sp_his5/CAN1 lyp1D/LYP1 his3-1/his3del1 leu2D0/LEU2 ura3D0/URA met5D0/MET5 LYS2/lys?? pCLB2::NatMX-MSH2 pCLB2::KanMX-SGS1	*S. uvarum* x *S. cerevisiae*
YTW165	Mata/Matα	can1D::STE2-Sp_his5/CAN1 lyp1D/LYP1 his3-1/his3del1 leu2D0/LEU2 ura3D0/URA met5D0/MET5 LYS2/lys?? pCLB2::NatMX-MSH2 pCLB2::KanMX-SGS1	*S. uvarum* x *S. cerevisiae*
R1158 *NET1*	Mata	URA::CMV-tTA kanR-tetO7-TATA–*NET1*	*S. cerevisiae*

		his3-1 leu2-0met15-0	
R1158 *MYO2*	Mata	URA::CMV-tTA kanR-tetO7-TATA–*MYO2* his3-1 leu2-0met15-0	*S. cerevisiae*
R1158 *MCM10*	Mata	URA::CMV-tTA kanR-tetO7-TATA–*MCM10* his3-1 leu2-0met15-0	*S. cerevisiae*
R1158 *CDC25*	Mata	URA::CMV-tTA kanR-tetO7-TATA–*CDC25* his3-1 leu2-0met15-0	*S. cerevisiae*
ed5652	Mata/Matα	URA3::Z3pr-YLR310C HAP1+::NatMX::pACT1-Z3EV-ENO2term/HAP1+ ura3^0/ura3^0 can1^Ste2pr-Sphis5/CAN1 his3^1/his3^1 lyp^0/LYP1	*S. cerevisiae*
ed4997	Mata/Matα	URA3::Z3pr-YIL109C HAP1+::NatMX::pACT1-Z3EV-ENO2term/HAP1+ ura3^0/ura3^0 can1^Ste2pr-Sphis5/CAN1 his3^1/his3^1 lyp^0/LYP1	*S. cerevisiae*
ed5440	Mata/Matα	URA3::Z3pr-YIL150C HAP1+::NatMX::pACT1-Z3EV-ENO2term/HAP1+ ura3^0/ura3^0 can1^Ste2pr-Sphis5/CAN1 his3^1/his3^1 lyp^0/LYP1	*S. cerevisiae*
ed5762	Mata/Matα	URA3::Z3pr-YNL152W HAP1+::NatMX::pACT1-Z3EV-ENO2term/HAP1+ ura3^0/ura3^0 can1^Ste2pr-Sphis5/CAN1 his3^1/his3^1 lyp^0/LYP1	*S. cerevisiae*
ed5373	Mata/Matα	URA3::Z3pr-YJL012C HAP1+::NatMX::pACT1-Z3EV-ENO2term/HAP1+ ura3^0/ura3^0	*S. cerevisiae*

		can1^Ste2pr-Sphis5/CAN1 his3^1/his3^1 lyp^0/LYP1	
ed5020	Mata/Matα	URA3::Z3pr-YDL212W HAP1+::NatMX::pACT1-Z3EV-ENO2term/HAP1+ ura3^0/ura3^0 can1^Ste2pr-Sphis5/CAN1 his3^1/his3^1 lyp^0/LYP1	*S. cerevisiae*
ed5538	Mata/Matα	URA3::Z3pr-YAL001C HAP1+::NatMX::pACT1-Z3EV-ENO2term/HAP1+ ura3^0/ura3^0 can1^Ste2pr-Sphis5/CAN1 his3^1/his3^1 lyp^0/LYP1	*S. cerevisiae*
ed4852	Mata/Matα	URA3::Z3pr-YOR326W HAP1+::NatMX::pACT1-Z3EV-ENO2term/HAP1+ ura3^0/ura3^0 can1^Ste2pr-Sphis5/CAN1 his3^1/his3^1 lyp^0/LYP1	*S. cerevisiae*
YTW260	Mata	myo2delta::hygMX his3 LYS2? ura3	*S. uvarum*
YTW225	Mata	*sec24Δ::HygMX HIS3 ura3 lys2*	*S. uvarum*
YTW253	Mata	tfc3delta::hygMX his3 LYS2? ura3	*S. uvarum*
YTW251	Matα	ura3Δ0 leu2Δ0 his3Δ1 lys2? Met15? Can1? myo2delta::NatMX with MoBY (CEN) rescue plasmid	*S. cerevisiae*
YTW221	Matα	ura3Δ0 leu2Δ0 his3Δ1 lys2? Met15? Can1? sec24delta::NatMX with MoBY (CEN) rescue plasmid	*S. cerevisiae*
YTW223	Matα	ura3Δ0 leu2Δ0 his3Δ1 lys2? Met15? Can1? tfc3delta::NatMX with MoBY (CEN) rescue plasmid	*S. cerevisiae*

YTW233	Mata/Matα	ura3Δ0/ura3 leu2Δ0/LEU2 his3Δ1/HIS3 lys2?/lys2 Met15?/MET15 Can1?/CAN1 sec24Δ::natMX/sec24Δ::hygMX	*S. uvarum* x *S. cerevisiae*
YTW243	Mata/Matα	can1D::STE2-Sp_his5/CAN1 lyp1D/LYP1 his3-delta1::CRE/HIS2 leu2D0/LEU2 ura3delta::CRE/ura3D0 met5D0/MET5 LYS2/lys2	*S. uvarum* x *S. cerevisiae*
YTW255	Mata/Matα	ura3Δ0/ura3 leu2Δ0/LEU2 his3Δ1/HIS3 lys2?/lys2 Met15?/MET15 Can1?/CAN1 tfc3Δ::natMX/tfc3Δ::hygMX	*S. uvarum* x *S. cerevisiae*
YTW263	Mata/Matα	ura3Δ0/ura3 leu2Δ0/LEU2 his3Δ1/HIS3 lys2?/lys2 Met15?/MET15 Can1?/CAN1 myo2Δ::natMX/myo2Δ::hygMX	*S. uvarum* x *S. cerevisiae*

Primers

Name	Sequence (5’ -> 3’)
TW103 - InvR transposon Sanger	GATGTCCTAAATGCACAGCGAC
TW104 - InvF transposon Sanger	GAGGCGTGCTTGTCAATGC
TW196 - Mod2_Tru Seq	Phos-GATCGGAAGAGCACACGTCTGAACTCCA GTCA
TW197 - Mod2_TS_ Univ	ACGCTCTTCCGATC*T

TW195 - FRH-seq library primer (P5)	ATGATACGGCGACCACCGAGATCTACACTCTT TCCCTACACGACGCTCTTCCGATCTGTCGTA AGCAATATTTCAAGAATGCATGCGTCAAT
TW198 - R amp transposon primer [adaptor] - index 1	CAAGCAGAAGACGGCATACGAGATGCTCAAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT
TW199 - R amp transposon primer [adaptor] - index 2	CAAGCAGAAGACGGCATACGAGATTCATGCGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT
TW200 - R amp transposon primer [adaptor] - index 3	CAAGCAGAAGACGGCATACGAGATATCGTAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT
TW201 - R amp transposon primer [adaptor] - index 4	CAAGCAGAAGACGGCATACGAGATCCAGTGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT
TW202 - R amp transposon primer [adaptor] - index 4	CAAGCAGAAGACGGCATACGAGATCTAGGAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT
TW203 - R amp	CAAGCAGAAGACGGCATACGAGATCGTAGGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCT

transposon primer [adaptor] - index 5	
TW123 - R amp transposon primer [adaptor] - index 6	CAAGCAGAAGACGGCATACGAGATTTGGTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TW124 - R amp transposon primer [adaptor] - index 7	CAAGCAGAAGACGGCATACGAGATCGCCCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
TW125 - R amp transposon primer [adaptor] - index 8	CAAGCAGAAGACGGCATACGAGATGCAACCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

References

1. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, Danford T, Bernstein DA, Alexander Rolfe P, Heisler LE, Chin B, Nislow C, Giaever G, Phillips PC, Fink GR, Gifford DK, Boone C. Genotype to Phenotype: A Complex Problem. Science. 2010 Apr;328(5977):425–425.

2. Gould TDF by SJ. Genetics and the Origin of Species: Columbia Classics edition. Columbia University Press; 1982. p. 364 Pages.

3. van Dijk PJ, Weissing FJ, Ellis THN. How Mendel's Interest in Inheritance Grew out of Plant Improvement. Genetics. 2018 Oct;210(2):347–355. PMCID: PMC6216597

4. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. Nat Methods. Nature Publishing Group; 2014 Aug;11(8):801–807.

5. Lee JM, Huddleston J, Doud MB, Hooper KA, Wu NC, Bedford T, Bloom JD. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 2018 Aug 28;115(35):E8276–E8285.

6. Spencer JM, Zhang X. Deep mutational scanning of S. pyogenes Cas9 reveals important functional domains. Sci Rep. Nature Publishing Group; 2017 Dec 4;7(1):16836.

7. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, Bloom JD. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell. 2020 Sep;182(5):1295-1310.e20.

8. Jones EM, Lubock NB, Venkatakrishnan A, Wang J, Tseng AM, Paggi JM, Latorraca NR, Cancilla D, Satyadi M, Davis JE, Babu MM, Dror RO, Kosuri S. Structural and functional characterization of G protein–coupled receptors with deep mutational scanning. Larhammar D, Aldrich RW, Fraser JS, Manglik A, editors. eLife. eLife Sciences Publications, Ltd; 2020 Oct 21;9:e54895.

9. Schott D, Ho J, Pruyne D, Bretscher A. The COOH-terminal domain of Myo2p, a yeast myosin V, has a direct role in secretory vesicle targeting. J Cell Biol. 1999 Nov 15;147(4):791–808. PMCID: PMC2156159

10. Govindan B, Bowser R, Novick P. The role of Myo2, a yeast class V myosin, in vesicular transport. J Cell Biol. 1995 Mar 15;128(6):1055–1068.

11. Lillie SH, Brown SS. Immunofluorescence localization of the unconventional myosin, Myo2p, and the putative kinesin-related protein, Smy1p, to the same regions of polarized growth in Saccharomyces cerevisiae. J Cell Biol. 1994 May 15;125(4):825–842.

12. Wendeler MW, Paccaud JP, Hauri HP. Role of Sec24 isoforms in selective export of membrane proteins from the endoplasmic reticulum. EMBO Rep. John Wiley & Sons, Ltd; 2007 Mar;8(3):258–264.

13. Lefebvre O, Carles C, Conesa C, Swanson RN, Bouet F, Riva M, Sentenac A. TFC3: gene encoding the B-block binding subunit of the yeast transcription factor IIIC. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 1992 Nov;89(21):10512–10516.

14. Carlborg Ö, Haley CS. Epistasis: too often neglected in complex trait studies? Nat Rev Genet. Nature Publishing Group; 2004 Aug;5(8):618–625.

15. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Hin Yan Tong A, van Dyk N, Wallace IM, Whitney JA, Weirauch MT, Zhong G, Zhu H, Houry WA, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser CA, Myers CL, Andrews BJ, Boone C. The Genetic Landscape of a Cell [Internet]. Available from: http://science.sciencemag.org/

16. Hin A, Tong Y, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Me P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu A marie, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C. Global Mapping of the Yeast Genetic Interaction Network. Science. 2004 Feb;303(5659):808–814.

17. Kuzmin E, VanderSluis B, Wang W, Tan G, Deshpande R, Chen Y, Usaj M, Balint A, Mattiazzi Usaj M, van Leeuwen J, Koch EN, Pons C, Dagilis AJ, Pryszlak M, Wang JZY, Hanchard J, Riggi M, Xu K, Heydari H, San Luis BJ, Shuteriqi E, Zhu H, Van Dyk N, Sharifpoor S, Costanzo M, Loewith R, Caudy A, Bolnick D, Brown GW, Andrews BJ, Boone C, Myers CL. Systematic analysis of complex genetic interactions. Science. 2018 Apr 20;360(6386):eaao1729.

18. Costanzo M, Vandersluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, Pelechano V, Styles EB, Billmann M, Van Leeuwen J, Van Dyk N, Lin ZY, Kuzmin E, Nelson J, Piotrowski JS, Srikumar T, Bahr S, Chen Y, Deshpande R, Kurat CF, Li SC, Li Z, Usaj M, Okada H, Pascoe N, Luis BJS, Sharifpoor S, Shuteriqi E, Simpkins SW, Snider J, Suresh HG, Tan Y, Zhu H, Malod-Dognin N, Janjic V, Przulj N, Troyanskaya OG, Stagljar I, Xia T, Ohya Y, Gingras AC, Raught B, Boutros M, Steinmetz LM, Moore CL, Rosebrock AP, Caudy AA, Myers CL, Andrews B, Boone C. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016 Sep;353(6306):aaf1420.

19. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A. Reverse engineering of regulatory networks in human B cells. Nat Genet. 2005 Apr;37(4):382–390.

20. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature. 2010 Apr;464(7291):1039–1042.

21. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, Qu H, Shales M, Park HO, Hayles J, Hoe KL, Kim DU, Ideker T, Grewal SI, Weissman JS, Krogan NJ. Conservation and Rewiring of Functional Modules Revealed by an Epistasis Map in Fission Yeast. Science. 2008 Oct;322(5900):405–410.

22. Hart GT, Lee I, Marcotte ER. A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality. 2007; Available from: http://www.biomedcentral.com/1471-2105/8/236

23. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001 May;411(6833):41–42.

24. Bader GD, Heilbut A, Andrews B, Tyers M, Hughes T, Boone C. Functional genomics and proteomics: charting a multidimensional map of the yeast cell. Trends Cell Biol. 2003 Jul;344–56.

25. Tarailo M, Tarailo S, Rose AM. Note Synthetic Lethal Interactions Identify Phenotypic "'Interologs'" of the Spindle Assembly Checkpoint Components. Available from: http://www.genetics.

26. Hartwell LH, Szankasi P, Roberts CJ, Murray AW, Friend SH. Integrating Genetic Approaches into the Discovery of Anticancer Drugs. Science. 1997;278(5340):1064–1068.

27. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA. 2009 Jun;9362–9367.

28. Katsanis N, Lupski JR, Beales PL. Exploring the molecular basis of Bardet-Biedl syndrome. Human Molecular Genetics. 2001 p. 2293–2299.

29. Zotenko E, Mestre J, O'leary DP, Przytycka TM. Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality. PLoS Comput Biol. 2008 Aug;4(8):e1000140.

30. Blomen VA, Májek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, Sacco R, Van Diemen FR, Olk N, Stukalov A, Marceau C, Janssen H, Carette JE, Bennett KL, Colinge J, Superti-Furga G, Brummelkamp TR. Gene essentiality and synthetic lethality in haploid human cells. Science. American Association for the Advancement of Science; 2015 Nov;350(6264):1092–1096.

31. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C,

Burtt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN. Common variants at 30 loci contribute to polygenic dyslipidemia. Marju Orho-Melander. 2009 Jan;18:56–65.

32. Jackson CA, Castro DM, Saldi GA, Bonneau R, Gresham D. Gene regulatory network reconstruction using single-cell RNA sequencing of barcoded genotypes in diverse environments. eLife. 2020;9:1–34.

33. Carette JE, Guimaraes CP, Wuethrich I, Blomen VA, Varadarajan M, Sun C, Bell G, Yuan B, Muellner MK, Nijman SM, Ploegh HL, Brummelkamp TR. Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. Nat Biotechnol. 2011 Jun;29(6):542–546.

34. Byrne AB, Weirauch MT, Wong V, Koeva M, Dixon SJ, Stuart JM, Roy PJ. A global analysis of genetic interactions in Caenorhabditis elegans. J Biol. 2007 Aug;6(3).

35. Aravind L, Watanabe H, Lipman DJ, Koonin EV. Lineage-specific loss and divergence of functionally linked genes in eukaryotes [Internet]. Available from: www.pnas.org

36. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, Welchman DP, Zipperien P, Ahringer J. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature. 2003 Jan;421(6920):231–237.

37. Dixon SJ, Fedyshyn Y, Koh JLY, Keshava Prasad TS, Chahwan C, Chua G, Toufighi K, Baryshnikova A, Hayles J, Hoe KL, Kim DU, Park HO, Myers CL, Pandey A, Durocher D, Andrews BJ, Boone C, Rine J, Designed CB, Performed GC. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. PNAS. 2008;16653–16658.

38. Davierwala AP, Haynes J, Li Z, Brost RL, Robinson MD, Yu L, Mnaimneh S, Ding H, Zhu H, Chen Y, Cheng X, Brown GW, Boone C, Andrews BJ, Hughes TR. The synthetic genetic interaction spectrum of essential genes. Nat Genet. 2005;37(10):1147–1152.

39. Dujon B. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. Trends Genet. Elsevier; 2006 Jul 1;22(7):375–387. PMID: 16730849

40. Botstein D, Fink GR. Yeast: An Experimental Organism for 21st Century Biology. Genetics. 2011 Nov;189(3):695–704.

41. Botstein D, Fink GR. Yeast: an experimental organism for modern biology. Science. 1988;240(4858):1439–1443.

42. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, Chu AM, Connelly C, Davis K, Dietrich F, Dow SW, El Bakkoury M, Foury F, Friend SH, Gentalen E, Giaever G, Hegemann JH, Jones T, Laub M, Liao H, Liebundguth N, Lockhart DJ, Lucau-Danila A, Lussier M, M'Rabet N, Menard P, Mittmann

M, Pai C, Rebischung C, Revuelta JL, Riles L, Roberts CJ, Ross-MacDonald P, Scherens B, Snyder M, Sookhai-Mahadeo S, Storms RK, Véronneau S, Voet M, Volckaert G, Ward TR, Wysocki R, Yen GS, Yu K, Zimmermann K, Philippsen P, Johnston M, Davis RW. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science. 1999;285(5429):901–906.

43. Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, Hieter P, Spencer F, Boeke JD. A Robust Toolkit for Functional Profiling of the Yeast Genome. Mol Cell. 2004 Nov;16(3):487–496.

44. Hilton SK, Huddleston J, Black A, North K, Dingens AS, Bedford T, Bloom JD. *dms-view*: Interactive visualization tool for deep mutational scanning data. J Open Source Softw. 2020 Aug 17;5(52):2353.

45. Melamed D, Young DL, Gamble CE, Miller CR, Fields S. Deep mutational scanning of an RRM domain of the Saccharomyces cerevisiae poly(A)-binding protein. RNA. 2013 Nov 1;19(11):1537–1551.

46. Hou J, Tan G, Fink GR, Andrews BJ, Boone C. Complex modifier landscape underlying genetic background effects. Proc Natl Acad Sci U S A. 2019;116(11):5045–5054.

47. Goldstein I, Ehrenreich IM. The complex role of genetic background in shaping the effects of spontaneous and induced mutations. Yeast. 2021;38(3):187–196.

48. Chow CY. Bringing genetic background into focus. Nat Rev Genet. 2016 Feb;17(2):63–64.

49. Chandler CH, Chari S, Dworkin I. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet. 2013 Jun 1;29(6):358–366.

50. Mullis MN, Matsui T, Schell R, Foree R, Ehrenreich IM. The complex underpinnings of genetic background effects. Nat Commun. 2018 Sep 17;9(1):3548.

51. All of Us Research Program | National Institutes of Health (NIH) [Internet]. All of Us Research Program | NIH. 2020 [cited 2023 Oct 23]. Available from: https://allofus.nih.gov/future-health-begins-all-us

52. Bloom JS, Ehrenreich IM, Loo WT, Lite TLV, Kruglyak L. Finding the sources of missing heritability in a yeast cross. Nature. Nature Publishing Group; 2013 Feb;494(7436):234–237.

53. Chin BL, Ryan O, Lewitter F, Boone C, Fink GR. Genetic Variation in Saccharomyces cerevisiae: Circuit Diversification in a Signal Transduction Network. Genetics. 2012 Dec 1;192(4):1523–1532.

54. Caudal E, Friedrich A, Jallet A, Garin M, Hou J, Schacherer J. Loss-of-function mutation survey revealed that genes with background-dependent fitness are rare and functionally related in yeast. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 2022 Sep 13;119(37):e2204206119.

55. Sanchez MR, Payen C, Cheong F, Hovde BT, Bissonnette S, Arkin AP, Skerker JM, Brem RB, Caudy AA, Dunham MJ. Transposon insertional mutagenesis in Saccharomyces uvarum reveals trans-acting effects influencing species-dependent essential genes. Genome Res. 2019;29(3):396–406.

56. Pulvirenti A, Nguyen HV, Caggia C, Giudici P, Rainieri S, Zambonelli C. Saccharomyces uvarum, a proper species within Saccharomyces sensu stricto. FEMS Microbiol Lett. 2000;192(2):191–196.

57. Nguyen HV, Boekhout T. Characterization of Saccharomyces uvarum (Beijerinck, 1898) and related hybrids: assessment of molecular markers that predict the parent and hybrid genomes and a proposal to name yeast hybrids. FEMS Yeast Res. 2017 Mar 1;17(2):fox014.

58. Almeida P, Gonc¸alves C, Teixeira S, Libkind D, Bontrager M, Masneuf-Pomarède I, Albertin W, Durrens P, Sherman DJ, Marullo P, Hittinger CT, Gonc¸alves P, Sampaio JP. A Gondwanan imprint on global diversity and domestication of wine and cider yeast Saccharomyces uvarum. Nat Commun [Internet]. 2014; Available from: www.nature.com/naturecommunications

59. Caudy AA, Guan Y, Jia Y, Hansen C, DeSevo C, Hayes AP, Agee J, Alvarez-Dominguez JR, Arellano H, Barrett D, Bauerle C, Bisaria N, Bradley PH, Scott Breunig J, Bush E, Cappel D, Capra E, Chen W, Clore J, Combs PA, Doucette C, Demuren O, Fellowes P, Freeman S, Frenkel E, Gadala-Maria D, Gawande R, Glass D, Grossberg S, Gupta A, Hammonds-Odie L, Hoisos A, Hsi J, Hsu YHH, Inukai S, Karczewski KJ, Ke X, Kojima M, Leachman S, Lieber D, Liebowitz A, Liu J, Liu Y, Martin T, Mena J, Mendoza R, Myhrvold C, Millian C, Pfau S, Raj S, Rich M, Rokicki J, Rounds W, Salazar M, Salesi M, Sharma R, Silverman S, Singer C, Sinha S, Staller M, Stern P, Tang H, Weeks S, Weidmann M, Wolf A, Young C, Yuan J, Crutchfield C, McClean M, Murphy CT, Llinás M, Botstein D, Troyanskaya OG, Dunham MJ. A new system for comparative functional genomics of Saccharomyces yeasts. Genetics. 2013;195(1):275–287.

60. Rolland T, Dujon B. Yeasty clocks: Dating genomic changes in yeasts. Comptes Rendus - Biol. 2011 Aug;334(8–9):620–628.

61. Brauer MJ, Christianson CM, Pai DA, Dunham MJ. Mapping Novel Traits by Array-Assisted Bulk Segregant Analysis in Saccharomyces cerevisiae. Genetics. 2006 Jul 1;173(3):1813–1816.

62. Dunham MJ. Two flavors of bulk segregant analysis in yeast. Methods Mol Biol. 2012;871:41–54. PMID: 22565832

63. Magwene PM, Willis JH, Kelly JK. The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing. PLoS Comput Biol. 2011;7(11):1002255–1002255.

64. de Barros Lopes M, Bellon JR, Shirley NJ, Ganter PF. Evidence for multiple interspecific hybridization in Saccharomyces sensu stricto species. FEMS Yeast Res. 2002 Jan;1(4):323–331.

65. Fischer G, James SA, Roberts IN, Oliver SG, Louis EJ. Chromosomal evolution in Saccharomyces. Nature. Nature Publishing Group; 2000 May;405(6785):451–454.

66. Gabaldón T. Hybridization and the origin of new yeast lineages. FEMS Yeast Res. 2020 Aug 1;20(5):foaa040.

67. Dujon BA, Louis EJ. Genome Diversity and Evolution in the Budding Yeasts (Saccharomycotina). Genetics. 2017 Jun 1;206(2):717–750.

68. Smukowski Heil CS, DeSevo CG, Pai DA, Tucker CM, Hoang ML, Dunham MJ. Loss of Heterozygosity Drives Adaptation in Hybrid Yeast. Mol Biol Evol. 2017 Jul;34(7):1596–1612.

69. Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. Two Dobzhansky-Muller Genes Interact to Cause Hybrid Lethality in Drosophila. Science. American Association for the Advancement of Science; 2006 Nov 24;314(5803):1292–1295.

70. A Rapidly Evolving Homeobox at the Site of a Hybrid Sterility Gene | Science [Internet]. [cited 2023 Oct 25]. Available from: https://www.science.org/doi/10.1126/science.282.5393.1501

71. Oka A, Mita A, Sakurai-Yamatani N, Yamamoto H, Takagi N, Takano-Shimizu T, Toshimori K, Moriwaki K, Shiroishi T. Hybrid breakdown caused by substitution of the X chromosome between two mouse subspecies. Genetics. 2004 Feb;166(2):913–924. PMCID: PMC1470736

72. Coughlan JM, Matute DR. The importance of intrinsic postzygotic barriers throughout the speciation process. Philos Trans R Soc B Biol Sci. Royal Society; 2020 Jul 13;375(1806):20190533.

73. Ferriol M, Simó U, Mansanet CJ, Torres A, Picó B, Monforte AJ, Romero C. Pre- and Post-Zygotic Barriers Contribute to Reproductive Isolation and Correlate with Genetic Distance in Cucumis. Plants Basel Switz. 2023 Feb 17;12(4):926. PMCID: PMC9963866

74. Monotilla AP, Nishimura T, Adachi M, Tanii Y, Largo DB, Hiraoka M. Examination of prezygotic and postzygotic isolating barriers in tropical Ulva (Ulvophyceae, Chlorophyta): evidence for ongoing speciation. J Phycol. 2018 Aug;54(4):539–549. PMID: 29846951

75. Turissini DA, McGirr JA, Patel SS, David JR, Matute DR. The Rate of Evolution of Postmating-Prezygotic Reproductive Isolation in Drosophila. Mol Biol Evol. 2018 Feb 1;35(2):312–334. PMCID: PMC5850467

76. Delneri D, Colson I, Grammenoudi S, Roberts IN, Louis EJ, Oliver SG. Engineering evolution to study speciation in yeasts. Nature. Nature Publishing Group; 2003 Mar;422(6927):68–72.

77. Hunter N, Chambers SR, Louis EJ, Borts RH. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. EMBO J. 1996;15(7):1726–1733.

78. Amin AD, Chaix ABH, Mason RP, Badge RM, Borts RH. The roles of the Saccharomyces cerevisiae RecQ helicase SGS1in meiotic genome surveillance. PLoS ONE. 2010;5(11).

79. Bozdag GO, Ono J, Denton JA, Karakoc E, Hunter N, Leu JY, Greig D. Engineering recombination between diverged yeast species reveals genetic incompatibilities. bioRxiv. 2019 Sep;755165–755165.

80. Carlile TM, Amon A. Meiosis I is established through division-specific translational control of a cyclin. Cell. 2008 Apr 18;133(2):280–291. PMCID: PMC2396536

81. Maher M, Cong F, Kindelberger D, Nasmyth K, Dalton S. Cell cycle-regulated transcription of the CLB2 gene is dependent on Mcm1 and a ternary complex factor. Mol Cell Biol. 1995 Jun;15(6):3129–3137. PMCID: PMC230544

82. Transposons | Learn Science at Scitable [Internet]. [cited 2023 Oct 5]. Available from: http://www.nature.com/scitable/topicpage/transposons-the-jumping-genes-518

83. Zhao S, Jiang E, Chen S, Gu Y, Shangguan AJ, Lv T, Luo L, Yu Z. PiggyBac transposon vectors: the tools of the human gene encoding. Transl Lung Cancer Res. 2016 Feb;5(1):120–125. PMCID: PMC4758974

84. de Lorenzo V, Timmis KN. [31] Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. Methods Enzymol [Internet]. Academic Press; 1994 [cited 2023 Oct 17]. p. 386–405. Available from: https://www.sciencedirect.com/science/article/pii/0076687994351570

85. Hellen EHB, Brookfield JFY. The Diversity of Class II Transposable Elements in Mammalian Genomes Has Arisen from Ancestral Phylogenetic Splits during Ancient Waves of Proliferation through the Genome. Mol Biol Evol. 2013 Jan;30(1):100–108. PMCID: PMC3525145

86. Nested Retrotransposons in the Intergenic Regions of the Maize Genome | Science [Internet]. [cited 2023 Oct 5]. Available from: https://www-science-org.offcampus.lib.washington.edu/doi/10.1126/science.274.5288.765

87. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. Nature. Nature Publishing Group; 1980 Apr;284(5757):601–603.

88. Orgel LE, Crick FHC. Selfish DNA: the ultimate parasite. Nature. Nature Publishing Group; 1980 Apr;284(5757):604–607.

89. NELSON PN, HOOLEY P, RODEN D, DAVARI EJTEHADI H, RYLANCE P, WARREN P, MARTIN J, MURRAY PG. Human endogenous retroviruses: transposable elements with potential ? Clin Exp Immunol. 2004 Oct;138(1):1–9. PMCID: PMC1809191

90. Carr M, Bensasson D, Bergman CM. Evolutionary Genomics of Transposable Elements in Saccharomyces cerevisiae. PLoS ONE. 2012 Nov 30;7(11):e50978. PMCID: PMC3511429

91. Coffin JM, Hughes SH, Varmus HE. Evolution of Retroelements and their Hosts. Retroviruses [Internet]. Cold Spring Harbor Laboratory Press; 1997 [cited 2023 Oct 26]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK19467/

92. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. Nature. Nature Publishing Group; 2004 Apr;428(6983):617–624.

93. Wolfe KH. Origin of the Yeast Whole-Genome Duplication. PLoS Biol. 2015 Aug 7;13(8):e1002221. PMCID: PMC4529243

94. Wolfe KH, Shields DC. Molecular evidence for an ancient duplication of the entire yeast genome. Nature. Nature Publishing Group; 1997 Jun;387(6634):708–713.

95. Adaptive genome duplication affects patterns of molecular evolution in Saccharomyces cerevisiae | PLOS Genetics [Internet]. [cited 2023 Oct 26]. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007396

96. Brideau NJ, Flores HA, Wang J, Maheshwari S, Wang X, Barbash DA. Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. Science. 2006 Nov 24;314(5803):1292–1295. PMID: 17124320

97. Fitzpatrick BM. Dobzhansky–Muller model of hybrid dysfunction supported by poor burst-speed performance in hybrid tiger salamanders. J Evol Biol. 2008;21(1):342–351.

98. Cutter AD. The polymorphic prelude to Bateson–Dobzhansky–Muller incompatibilities. Trends Ecol Evol. 2012 Apr 1;27(4):209–218.

99. Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JS (Pat). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Ann Bot. 2017 Aug 1;120(2):183–194.

100. Paun O, Forest F, Fay MF, Chase MW. Hybrid speciation in angiosperms: parental divergence drives ploidy. New Phytol. 2009;182(2):507–518. PMCID: PMC2988484

101. Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY. Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility between Two Yeast Species. Cell. Elsevier; 2008 Dec 12;135(6):1065–1073. PMID: 19070577

102. Fishel R. Mismatch Repair*. J Biol Chem. 2015 Oct 30;290(44):26395–26403.

103. Groothuizen FS, Sixma TK. The conserved molecular machinery in DNA mismatch repair enzyme structures. DNA Repair. 2016 Feb 1;38:14–23.

104. Pećina-Šlaus N, Kafka A, Salamon I, Bukovac A. Mismatch Repair Pathway, Genome Stability and Cancer. Front Mol Biosci [Internet]. 2020 [cited 2023 Oct 26];7. Available from: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00122

105. Spies M, Fishel R. Mismatch Repair during Homologous and Homeologous Recombination. Cold Spring Harb Perspect Biol. 2015 Mar 1;7(3):a022657.

106. Elez M. Mismatch Repair: From Preserving Genome Stability to Enabling Mutation Studies in Real-Time Single Cells. Cells. Multidisciplinary Digital Publishing Institute; 2021 Jun;10(6):1535.

107. Manhart CM, Alani E. Roles for mismatch repair family proteins in promoting meiotic crossing over. DNA Repair. 2016 Feb 1;38:84–93.

108. Goldfarb T, Alani E. Distinct roles for the Saccharomyces cerevisiae mismatch repair proteins in heteroduplex rejection, mismatch repair and nonhomologous tail removal. Genetics. 2005;169(2):563–574.

109. Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. Nature. Nature Publishing Group; 1989 Nov;342(6248):396–401.

110. Worth L, Clark S, Radman M, Modrich P. Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 1994 Apr 12;91(8):3238–3241.

111. Polosina YY, Mui J, Pitsikas P, Cupples CG. The Escherichia coli Mismatch Repair Protein MutL Recruits the Vsr and MutH Endonucleases in Response to DNA Damage. J Bacteriol. American Society for Microbiology; 2009 Jun 15;191(12):4041–4043.

112. Matic I, Rayssiguier C, Radman M. Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species. Cell. 1995 Feb 10;80(3):507–515.

113. New L, Liu K, Crouse GF. The yeast gene MSH3 defines a new class of eukaryotic MutS homologues. Mol Gen Genet MGG. 1993 May 1;239(1):97–108.

114. Borts RH, Leung WY, Kramer W, Kramer B, Williamson M, Fogel S, Haber JE. Mismatch repair-induced meiotic recombination requires the pms1 gene product. Genetics. 1990 Mar 1;124(3):573–584.

115. Datta A, Adjiri A, New L, Crouse GF, Jinks-Robertson S. Mitotic Crossovers between Diverged Sequences Are Regulated by Mismatch Repair Proteins in Saccharomyces cerevisiae. Mol Cell Biol. Taylor & Francis; 1996 Mar 1;16(3):1085–1093. PMID: 8622653

116. MLH1, PMS1, and MSH2 Interactions During the Initiation of DNA Mismatch Repair in Yeast | Science [Internet]. [cited 2023 Oct 26]. Available from: https://www.science.org/doi/10.1126/science.8066446

117. Alani E, Reenan RA, Kolodner RD. Interaction between mismatch repair and genetic recombination in Saccharomyces cerevisiae. Genetics. 1994 May 1;137(1):19–39.

118. Radman M, Wagner R. Mismatch recognition in chromosomal interactions and speciation. Chromosoma. 1993 Jun 1;102(6):369–373.

119. Greig D. Reproductive isolation in Saccharomyces. Heredity. Nature Publishing Group; 2009 Jan;102(1):39–44.

120. Grandin N, Reed SI. Differential function and expression of Saccharomyces cerevisiae B-type cyclins in mitosis and meiosis. Mol Cell Biol. 1993;13(4):2113–2125.

121. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. Nature Publishing Group; 2007 Jan;2(1):31–34.

122. Samsonova IA, Böttcher F, Steinwehr D, Schilowa B. A simple technique for random spore analysis of yeasts using nystatin. J Microbiol Methods. 1985 May 1;3(5):283–290.

123. Taxis C, Keller P, Kavagiou Z, Jensen LJ, Colombelli J, Bork P, Stelzer EHK, Knop M. Spore number control and breeding in Saccharomyces cerevisiae : a key role for a self-organizing system. J Cell Biol. 2005 Nov 14;171(4):627–640.

124. Haarer B, Aggeli D, Viggiano S, Burke DJ, Amberg DC. Novel Interactions between Actin and the Proteasome Revealed by Complex Haploinsufficiency. PLOS Genet. Public Library of Science; 2011 Sep 22;7(9):e1002288.

125. Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CWV, Bussey H, Andrews B, Tyers M, Boone C. Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science. American Association for the Advancement of Science; 2001 Dec 14;294(5550):2364–2368.

126. Amorosi CJ, Chiasson MA, McDonald MG, Wong LH, Sitko KA, Boyle G, Kowalski JP, Rettie AE, Fowler DM, Dunham MJ. Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. Am J Hum Genet. Elsevier; 2021 Sep 2;108(9):1735–1751. PMID: 34314704

127. Zhang L, Sarangi V, Moon I, Yu J, Liu D, Devarajan S, Reid JM, Kalari KR, Wang L, Weinshilboum R. CYP2C9 and CYP2C19: Deep Mutational Scanning and Functional Characterization of Genomic Missense Variants. Clin Transl Sci. 2020;13(4):727–742.

128. Massively Parallel Functional Analysis of BRCA1 RING Domain Variants | Genetics | Oxford Academic [Internet]. [cited 2023 Oct 26]. Available from: https://academic.oup.com/genetics/article/200/2/413/5936195

129. Girgis HS, Liu Y, Ryu WS, Tavazoie S. A Comprehensive Genetic Characterization of Bacterial Motility. PLOS Genet. Public Library of Science; 2007 Sep 14;3(9):e154.

130. Meluh PB, Pan X, Yuan DS, Tiffany C, Chen O, Sookhai-Mahadeo S, Wang X, Peyser BD, Irizarry R, Spencer FA, Boeke JD. Analysis of genetic interactions on a genome-wide scale in budding yeast: diploid-based synthetic lethality analysis by microarray. Methods Mol Biol Clifton NJ. 2008;416:221–247. PMID: 18392971

131. Tischler J, Lehner B, Fraser AG. Evolutionary plasticity of genetic interaction networks. Nat Genet. 2008 Apr;40(4):390–391.

132. Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic Dissection of Transcriptional Regulation in Budding Yeast. Science. American Association for the Advancement of Science; 2002 Apr 26;296(5568):752–755.

133. Collins SR, Schuldiner M, Krogan NJ, Weissman JS. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol. 2006;7(7):63–63.

134. Malmberg RL, Mauricio R. QTL-based evidence for the role of epistasis in evolution. Genet Res. Cambridge University Press; 2005 Oct;86(2):89–95.

135. Bergman A, Siegal ML. Evolutionary capacitance as a general feature of complex gene networks. Nature. 2003 Jul;424(6948):549–552.

136. Guo Y, Park JM, Cui B, Humes E, Gangadharan S, Hung S, FitzGerald PC, Hoe KL, Grewal SIS, Craig NL, Levin HL. Integration Profiling of Gene Function With Dense Maps of Transposon Integration. Genetics. 2013 Oct 1;195(2):599–609.

137. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ, McAdams HH, Shapiro L. The essential genome of a bacterium. Mol Syst Biol. John Wiley & Sons, Ltd; 2011 Jan;7(1):528.

138. van Opijnen T, Bodi KL, Camilli A. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods. Nature Publishing Group; 2009 Oct;6(10):767–772.

139. Smith GA, Enquist LW. Construction and Transposon Mutagenesis inEscherichia coli of a Full-Length Infectious Clone of Pseudorabies Virus, an Alphaherpesvirus. J Virol. American Society for Microbiology; 1999 Aug;73(8):6405–6414.

140. Michel AH, Hatakeyama R, Kimmig P, Arter M, Peter M, Matos J, De Virgilio C, Kornmann BT. Functional mapping of yeast genomes by saturated transposition. eLife. 2017;6:1–28.

141. Weiss CV, Roop JI, Hackley RK, Chuong JN, Grigoriev IV, Arkin AP, Skerker JM, Brem RB. Genetic dissection of interspecific differences in yeast thermotolerance. Nat Genet. 2018;50(11):1501–1504.

142. van Leeuwen J, Pons C, Tan G, Wang JZ, Hou J, Weile J, Gebbia M, Liang W, Shuteriqi E, Li Z, Lopes M, Ušaj M, Dos Santos Lopes A, van Lieshout N, Myers CL, Roth FP, Aloy P, Andrews BJ, Boone C. Systematic analysis of bypass suppression of essential genes. Mol Syst Biol. John Wiley & Sons, Ltd; 2020 Sep;16(9):e9828.

143. Saccharomyces Genome Database | SGD [Internet]. [cited 2023 Oct 26]. Available from: https://www.yeastgenome.org/

144. Kleckner N, Roth J, Botstein D. Genetic engineering in Vivo using translocatable drug-resistance elements. New methods in bacterial genetics. J Mol Biol. 1977 Oct;116(1):125–159.

145. Serviene E, Jiang Y, Cheng CP, Baker J, Nagy PD. Screening of the Yeast yTHC Collection Identifies Essential Host Factors Affecting Tombusvirus RNA Recombination. J Virol. American Society for Microbiology; 2006 Feb;80(3):1231–1241.

146. Park YN, Morschhäuser J. Tetracycline-Inducible Gene Expression and Gene Deletion in Candida albicans. Eukaryot Cell. American Society for Microbiology; 2005 Aug;4(8):1328–1342.

147. McIsaac RS, Oakes BL, Wang X, Dummit KA, Botstein D, Noyes MB. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. Nucleic Acids Res. 2013 Feb 1;41(4):e57.

148. Zopf CJ, Quinn K, Zeidman J, Maheshri N. Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. PLOS Comput Biol. Public Library of Science; 2013 Jul 25;9(7):e1003161.

149. Tominaga M, Kondo A, Ishii J. Engineering of Synthetic Transcriptional Switches in Yeast. Life. Multidisciplinary Digital Publishing Institute; 2022 Apr;12(4):557.

150. Tominaga M, Nozaki K, Umeno D, Ishii J, Kondo A. Robust and flexible platform for directed evolution of yeast genetic switches. Nat Commun. Nature Publishing Group; 2021 Mar 23;12(1):1846.

151. Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet. Nature Publishing Group; 2007 Oct;8(10):762–775.

152. Harris S, Rudnicki KS, Haber JE. Gene conversions and crossing over during homologous and homeologous ectopic recombination in Saccharomyces cerevisiae. Genetics. 1993 Sep;135(1):5–16. PMCID: PMC1205626

153. Hawthorne D, Philippsen P. Genetic and molecular analysis of hybrids in the genus Saccharomyces involving S. cerevisiae, S. uvarum and a new species, S. douglasii. Yeast Chichester Engl. 1994 Oct;10(10):1285–1296. PMID: 7900417

154. Hsu YY, Chou JY. Environmental Factors Can Influence Mitochondrial Inheritance in the Saccharomyces Yeast Hybrids. PLOS ONE. Public Library of Science; 2017 Jan 12;12(1):e0169953.

155. Hewitt SK, Duangrattanalert K, Burgis T, Zeef LAH, Naseeb S, Delneri D. Plasticity of Mitochondrial DNA Inheritance and Its Impact on Nuclear Gene Transcription in Yeast Hybrids. Microorganisms. Multidisciplinary Digital Publishing Institute; 2020 Apr;8(4):494.

156. MoBY-ORF Molecular Barcoded Yeast ORF Library [Internet]. [cited 2023 Oct 26]. Available from: https://moby.ccbr.utoronto.ca/

157. Hoffman CS, Winston F. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformaion of Escherichia coli. Gene. 1987 Jan 1;57(2):267–272.

158. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A. Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. mBio. American Society for Microbiology; 2015 May 12;6(3):10.1128/mbio.00306-15.

159. Li X, Burnight ER, Cooney AL, Malani N, Brady T, Sander JD, Staber J, Wheelan SJ, Joung JK, McCray PB, Bushman FD, Sinn PL, Craig NL. piggyBac transposase tools for genome engineering. Proc Natl Acad Sci. Proceedings of the National Academy of Sciences; 2013 Jun 18;110(25):E2279–E2287.

160. Hedges SB. The origin and evolution of model organisms. Nat Rev Genet. Nature Publishing Group; 2002 Nov;3(11):838–849.

161. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science. American Association for the Advancement of Science; 2014 Nov 28;346(6213):1258096.

162. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity | Science [Internet]. [cited 2023 Oct 26]. Available from: https://www.science.org/doi/10.1126/science.1225829

163. Nong W, Cao J, Li Y, Qu Z, Sun J, Swale T, Yip HY, Qian PY, Qiu JW, Kwan HS, Bendena W, Tobe S, Chan TF, Yip KY, Chu KH, Ngai SM, Tsim KY, Holland PWH, Hui JHL. Jellyfish genomes reveal distinct homeobox gene clusters and conservation of small RNA processing. Nat Commun. Nature Publishing Group; 2020 Jun 19;11(1):3051.

164. Willey C, Korstanje R. Sequencing and assembling bear genomes: the bare necessities. Front Zool. 2022 Nov 30;19(1):30.

165. Recuerda M, Vizueta J, Cuevas-Caballé C, Blanco G, Rozas J, Milá B. Chromosome-Level Genome Assembly of the Common Chaffinch (Aves: Fringilla coelebs): A Valuable Resource for Evolutionary Biology. Genome Biol Evol. 2021 Apr 1;13(4):evab034.

166. Hyun JY, Pandey P, Kim KS, Chon A, Jeong D, Bhak J, Yu M, Song HK, Singh R, Min MS, Goyal SP, Bayarkhagva D, Marchenkova T, Vitkalova A, Lee H. Whole genome survey of big cats (Genus: Panthera) identifies novel microsatellites of utility in conservation genetic study. Sci Rep. Nature Publishing Group; 2021 Jul 8;11(1):14164.

167. Koutsovoulos GD, Poullet M, Elashry A, Kozlowski DKL, Sallet E, Da Rocha M, Perfus-Barbeoch L, Martin-Jimenez C, Frey JE, Ahrens CH, Kiewnick S, Danchin EGJ. Genome assembly and annotation of Meloidogyne enterolobii, an emerging parthenogenetic root-knot nematode. Sci Data. Nature Publishing Group; 2020 Oct 5;7(1):324.

168. Koochekian N, Ascanio A, Farleigh K, Card DC, Schield DR, Castoe TA, Jezkova T. A chromosome-level genome assembly and annotation of the desert horned lizard, Phrynosoma platyrhinos, provides insight into chromosomal rearrangements among reptiles. GigaScience. 2022 Jan 1;11:giab098.

169. The Human Genome Project [Internet]. Genome.gov. [cited 2023 Oct 23]. Available from: https://www.genome.gov/human-genome-project

170. Qin D. Next-generation sequencing and its clinical application. Cancer Biol Med. Cancer Biology & Medicine; 2019 Feb 1;16(1):4–10.

171. Marx V. Method of the year: long-read sequencing. Nat Methods. Nature Publishing Group; 2023 Jan;20(1):6–11.

172. Pop M. Genome assembly reborn: recent computational challenges. Brief Bioinform. 2009 Jul 1;10(4):354–366.

173. Gill R, Datta S, Datta S. A statistical framework for differential network analysis from microarray data [Internet]. 2010. Available from: http://www.somnathdatta.org/Supp/DNA.

174. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 2003 Nov;13(11):2498–2504. PMCID: PMC403769

175. Zhang M, Li Q, Yu D, Yao B, Guo W, Xie Y, Xiao G. GeNeCK: A web server for gene network construction and visualization. BMC Bioinformatics. BioMed Central Ltd.; 2019 Jan;20(1):12–12.

176. Livesey BJ, Marsh JA. Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. Mol Syst Biol. John Wiley & Sons, Ltd; 2020 Jul;16(7):e9380.

177. Schuhmacher LN, Husson Z, Smith ESJ. The naked mole-rat as an animal model in biomedical research: current perspectives. Open Access Anim Physiol. Dove Press; 2015 Aug 11;7:137–148.

178. Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, Marino SM, Sun X, Turanov AA, Yang P, Yim SH, Zhao X, Kasaikina MV, Stoletzki N, Peng C, Polak P, Xiong Z, Kiezun A, Zhu Y, Chen Y, Kryukov GV, Zhang Q, Peshkin L, Yang L, Bronson RT, Buffenstein R, Wang B, Han C, Li Q, Chen L, Zhao W, Sunyaev SR, Park TJ, Zhang G, Wang J, Gladyshev VN. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature. Nature Publishing Group; 2011 Nov;479(7372):223–227.