The Contributions of Single Alleles to the Biology of Interspecific Hybrid *Saccharomyces*

Samuel M. Lancaster

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Maitreya Dunham, Chair

Judit Villen

Josh Akey

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

The Contributions of Single Alleles to the Biology of Interspecific Hybrid *Saccharomyces*

Samuel Lancaster

Chair of the Supervisory Committee:
Maitreya Dunham
Associate Professor of Genome Sciences
Department of Genome Sciences

Interspecific hybrids play integral roles in society and are essential to industries worth hundreds of billions of dollars. They often have dominant genetic networks that allow them to be more fit than either parent, particularly in manmade settings. Hybrid *Saccharomyces* are amenable for scientific investigation because many hybrids are sterile, but *Saccharomyces* can grow indefinitely mitotically thereby bypassing this problem. Further, *Saccharomyces cerevisiae* is perhaps the best studied organism in the world, with countless tools to facilitate research. I sought to understand the genome-wide effects of single allele deletions on hybrids and determine their individual contributions to hybrid fitness. I mated the *S. cerevisiae* deletion collection, which contains a deletion strain for nearly every ORF, to WT *S. uvarum,* which is 20% diverged on the amino acid and nucleotide level. I also mated two *S. cerevisiae* collections with conditional knockouts of essential genes to WT *S. uvarum*. I then found the fitness of every deletion strain in three chemostat media, and the fitness of the essential genes knockouts on solid YPD. I found the hybrid deletion strains had more within condition and between condition variances than the purebreds, as well as large gene by genetic background effects. The essential genes had significantly less variance between conditions than the nonessential deletions, and significantly less gene by genetic background effects. To investigate the effects on fitness of each ortholog, I performed a reciprocal hemizygosity analysis showing that the

effects of single allele deletions are dependent on the ortholog deleted. Together, our results show why hybrids are so abundant in manmade settings. They begin heterotic and they have new ways of exploring fitness landscape that have larger effects. To determine the molecular interactions between hybrid alleles, we developed a method, in conjunction with the Yates lab at Scripps, finding the preferences of particular orthologs in hybrid protein complexes, and showed that key amino acids can dramatically affect interspecific binding. Our findings show that hybrid proteins often have heterogeneous preferences for orthologs, which have been shown to confer heterosis. Lastly, I determined whether there are genes essential only to interspecific mating by mating the deletion collection to *S. uvarum,* and generating mating scores for every strain. Our findings did not identify any complete hybrid-specific mating incompatibilities, though there were significant incompatibilities suggesting it is possible to create a *S. cerevisiae* strain that does not mate with *S. uvarum* but that does mate with itself.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Any person who showed kindness on a bad day deserves some sort of acknowledgement for accomplishing this work. In particular, I want to thank the Dunham Lab for all their support. They saw me at my best and at my worst and have been supportive through it all. I have also had the opportunity to work with some incredibly intelligent collaborators who filled me with outstanding ideas that I could expand upon. No work such as this could possibly be accomplished alone, and I'm grateful to everyone who helped me.

DEDICATION

I dedicate this to my wife, Jessica Lancaster, who has changed my life in countless ways, all of them positively.

CHAPTER 1: INTRODUCTION: THE EFFECTS OF HYBRIDIZATION ON BIOLOGY AND ITS ROLE IN STUDYING EVOLUTION

Deleterious recessive alleles are widespread within any population. Highlighting the impact of these alleles are their effects on inbred populations, where they become homozygous. Such populations almost universally display rare diseases, or, as in single-celled organisms, have lower fitness than heterozygous populations. Conversely, outbreeding can mitigate the effects of these recessive alleles by dominant complementation. An extreme form of outbreeding is interspecific hybridization where two diverged species form an organism, which can have higher fitness than either parental species, known as hybrid vigor or heterosis. In metazoans, interspecific hybrids are relatively rare and often sterile. However, in virtually every other clade of life hybridization is more common, and hybrid plants and fungi are abundant and socially important. Wheat for example, a pillar of Western society, is a triple hybrid that creates more grain than any of its parental grass species (IWGSC et al., 2014). For these reasons, studying hybrids and the forces that contribute to heterosis is of paramount importance.

SACCHAROMYCES CEREVISIAE *IS AN ABUNDANT BUDDING YEAST AND CAN HYBRIDIZE WITH MANY OTHER SPECIES*

S. cerevisiae is a budding yeast, and part of the fungal clade. It is found ubiquitously around the world, and recently it has been suggested that it is a nomadic yeast adapted to multiple ecological niches (Goddard & Greig, 2015). Its most common ecological niches are places where it can ferment carbohydrates, including fruits and the guts of insects, which were likely its main vector for colonizing new areas before humans (Liti, 2015). Being ubiquitous, *S.*

cerevisiae has also been found from soil to tree bark, among other places (Liti, 2015). Taking advantage of carbohydrates in its environment, the main metabolic process of yeast is fermentation even under aerobic conditions, producing ethanol that poisons many of its competitors.

The unique properties of *S. cerevisiae* means it has been utilized by humans for producing food and drink for millennia. The first bread made 10,000 - 30,000 years ago may have been naturally leavened by *S. cerevisiae* spores (McGee, 2007), and starter yeast cultures for leavening bread have been documented since antiquity (Tannahill, 1989). Alcoholic beverages around the globe are also made by *S. cerevisiae,* including beer, wine, and cider. Beer brewing possibly dates back to the Neolithic period, and is documented as early as the ancient Egyptians. Beer has also been suggested to contribute to multiple aspects of society by playing an important social role and sterilizing water (Hayden & Villeneuve, 2011). Today the modern alcohol industry is worth hundreds of billions of dollars, and *S. cerevisiae,* and its close relatives, do most of the fermenting.

S. cerevisiae is an excellent model organism with a fast doubling time, economic value, a compact genome, amenability to molecular biological techniques, and, though unicellular, shares properties with higher eukaryotes, such as sexual reproduction and cellular structure. These properties have made it perhaps the best-studied eukaryote. Often, when transformative biological technologies are developed, *S. cerevisiae* is the first eukaryote to which they are applied. For example *S. cerevisiae* was the first eukaryote to have its genome sequenced (Goffeau et al., 1996). The first genome-wide deletion collection was also made in *S. cerevisiae*. The collection came out around 2000, and it has nearly every open reading frame (ORF) deleted (Giaever et al., 2002; Winzeler et al., 1999). This collection has led to countless discoveries

about the functions of this vast array of genes, with greater than 1000 genome-wide studies performed to date (Giaever & Nislow, 2014). These studies tend to highlight gene by environment effects, as many deletions have different effects on cellular fitness in different environments. Before the deletion collection, the first large scale deletion screens were done in a more crude way, with random transposon insertions into whole yeast chromosomes. This work was the first to identify gene by environment effects at a genomic level, and highlights the rapid development of the yeast field (Hodgkin, Jackson, Mak, Martienssen, & Rowley, 1995; V. Smith, Botstein, & Brown, 1995). The deletion collection is just one example of the many powerful tools available to *S. cerevisiae* researchers.

S. cerevisiae tolerates hybridization extremely well and can mate with many diverse budding yeast, including some that are up to 20% diverged on the nucleotide and amino acid levels (CC, JE, & JD, 1953). The *Saccharomyces sensu stricto* are all the species that can mate with *S. cerevisiae* and each other, but no other organisms (Fischer, James, Roberts, Oliver, & Louis, 2000; Scannell et al., 2011). These species include: *Saccharomyces paradoxus, Saccharomyces mikatae, Saccharomyces kudriavzevii, Saccharomyces arboricola, Saccharomyces eubayanus,* and *Saccharomyces uvarum* (Figure 1). These yeast all have 16 chromosomes, and are the decedents of an archaic 8 chromosome yeast that underwent a whole genome duplication, perhaps as a result of a hybridization event (Dietrich et al., 2004; Kellis, Birren, & Lander, 2004; Marcet-Houben & Gabaldón, 2015; Scannell, Byrne, Gordon, Wong, & Wolfe, 2006; Scannell et al., 2007; Wapinski, Pfeffer, Friedman, & Regev, 2007; Wolfe & Shields, 1997). With so many options for hybridization partners, the *Saccharomyces sensu stricto* is emerging as a primary clade to study hybrids.

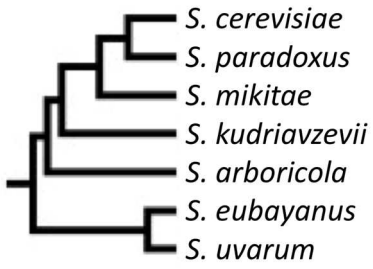

Figure 1.1. The *Saccharomyces sensu stricto*. (Hittinger, 2013)

Interspecific hybrids in the *sensu stricto* are common outside of the lab, fermenting many alcoholic beverages. Brewing environments are an excellent place for hybrids to form, as diverse yeast taken from geographically distant areas may closely interact, and they may be under strong, narrow, manmade selective pressures. Grain and hops taken from hundreds of miles apart, each with unique *Saccharomyces*, may mix in such environments. For example, two hybrid brewing yeasts have received their own designation as species – *Saccharomyces bayanus* and *Saccharomyces pastorianus*. These are combinations of *S. cerevisiae* and cryotolerant species – *S. uvarum* and *S. eubayanus* in the case of *S. bayanus*, and primarily *S. eubayanus* in the case of *S. pastorianus* – and have favorable fermentation qualities that often make better wines, beers, and ciders (Dunn & Sherlock, 2008; Libkind et al., 2011; Naumova, Naumov, Masneuf-Pomarède, Aigle, & Dubourdieu, 2005; Nguyen, Legras, Neuvéglise, & Gaillardin, 2011; Pérez-Través, Lopes, Querol, & Barrio, 2014). Hence, not only is *S. cerevisiae* able to mate with many other species, but it does so outside of the laboratory, and these hybrids share many properties with *S. cerevisiae,* including fast doubling time and economic importance.

INTERSPECIFIC HYBRIDS AS A MEANS TO STUDY SPECIATION

The widespread nature of hybridization, exemplified by both cereal crops and *Saccharomyces,* poses unique questions about speciation. At what point do two eukaryotic

organisms diverge enough so that they are genetically isolated via the methods of mating and meiosis? This problem has vexed biologists since Charles Darwin, who wondered, "Why, if species have descended from other species by fine gradations, do we not everywhere see innumerable transitional forms? Why is not all nature in confusion, instead of the species being, as we see them, well defined?*"* (Darwin, 1909) Though in yeast the gradations of phenotypes are by some measures less dramatic than in other eukaryotes, there are nonetheless boundaries of genetic flow between clades.

At the genetic level, hybrids tend to have interesting, aneuploid karyotypes relative to their euploid parents (Dunn & Sherlock, 2008; Nakao et al., 2009; Nguyen, Lepingle, & Gaillardin, 2000). Not only is it thought-provoking how these extensive heterogeneous aneuploidies contribute to heterosis, but these phenomena very directly relate to human well being. The most aggressive cancers, for example, are often highly aneuploid, and the reasons for this are not completely understood (D. J. Gordon, Resio, & Pellman, 2012). Perhaps hybrid *Saccharomyces* will yield valuable insights into this phenomenon.

In a study that evolved hybrid yeast through meiosis, the authors showed that hybrids shed entire chromosomes. The *S. cerevisiae* chromosomes were more frequently maintained (Antunovics, Nguyen, Gaillardin, & Sipiczki, 2005), perhaps because of incompatibilities between the *S. uvarum* genome and the *S. cerevisiae* mitochondria (Lee et al., 2008). Other studies evolving yeast have also observed drastic aneuploidies. One evolved hybrid yeast in chemostats with ammonium sulfate limitation. They found many different genome rearrangements, with no clear pattern for adopting a single chromosome except for a set of translocations which fused the *S. cerevisiae* copy of *MEP2* – an ammonium permease – with the *S. uvarum* copy (Dunn et al., 2013). These studies show that hybrids seem to be susceptible to

large genome arrangements that are often unpredictable, or that are driven by relatively narrow metabolic and genetic constraints.

Though the *sensu stricto* can all mate and form viable spores, *S. cerevisiae* mates poorly with *S. uvarum,* and only a few spores survive meiosis – around 0.01% (Sebastiani, Barberio, Casalone, Cavalieri, & Polsinelli, 2002). These two species have three reciprocal translocations between them (Naumov, Naumova, & Gaillardin, 1993; Nguyen et al., 2000; Scannell et al., 2011), which will reduce spore viability. Nonetheless, species of comparable divergence can form viable hybrid lineages, as is the case with *S. pastorianus*, albeit with significant aneuploidies (Dunn & Sherlock, 2008).

If organisms more genetically flexible than metazoans can tolerate, and even thrive with, significant aneuploidies, this begs the question: what in these organisms causes complete genetic isolation? When the mismatch repair system is disrupted in *S. cerevisiae* x *S. uvarum* hybrids spore viability increases to about 0.1%. In hybrids formed by more closely related *Saccharomyces* with no differing aneuploidies between them, there are still decreases in spore viability that correlate with divergence (Libkind et al., 2011). It has also been shown that translocation abundance does not correlate perfectly with post-zygotic isolation (Fischer et al., 2000). Together, these studies suggest that there are other mechanisms than translocations that increase, and may be required for, genetic isolation in *Saccharomyces*.

Nuclear-mitochondrial incompatibilities have been identified as a primary contributor to interspecies incompatibilities. One study found, consistent with what has been observed in our laboratory, that the hybrids always take the *S. cerevisiae* mitochondria (Chou, Hung, Lin, Lee, & Leu, 2010). Since laboratory conditions are 30°C, which are optimal for *S. cerevisiae* growth, perhaps the *S. cerevisiae* mitochondria is better suited to metabolism in this condition. *S.*

pastorianus, which ferments at cooler temperatures, primarily inherits the *S. eubayanus* mitochondria. Hence, the mitochondria inherited may be dependent on environment and can cause incompatibilities with portions of the hybrid genome.

This group finding *S. cerevisiae* mitochondria in hybrids also found that there is an incompatibility between the *S. uvarum* gene *AEP2* on chromosome 13 and the *S. cerevisiae* mitochondria because it cannot properly process *OLI1* mRNA from the *S. cerevisiae* mitochondrial chromosome (Lee et al., 2008). This leads to increased sterility of F1 hybrids. However, if the F1 obtains the *S. cerevisiae AEP2* or the *S. uvarum* mitochondria, and at least half of them will, this incompatibility will not exist. Further evidence for nuclear-mitochondrial incompatibilities came from the same group, which found the same two incompatibilities between *S. cerevisiae* and *S. uvarum* and between *S. cerevisiae* and *S. paradoxus*. These genomic incompatibilities act with the mitochondria in different ways – by affecting the splicing of mitochondrial *COX1* and mitochondrial lypolytic action. These show that nuclear-mitochondrial incompatibilities represent a generalized mechanism for reproductive isolation. Another study looking at *S. cerevisiae* x *S. uvarum* F1s found incompatibilities with the *S. uvarum MATa* locus on chromosome 2, which when eliminated, allowed the F1s to better undergo meiosis (Pfliegler, Antunovics, & Sipiczki, 2012). Together these studies have shown that interspecific incompatibilities are a generalized contributor to reproductive isolation; however, they do not cause complete isolation.

One way to generate complete reproductive isolation would be through prezygotic barriers to hybridization. Yeast, though unicellular, comes in two mating types, designated a and α. These types will emit pheromones and begin mating after pheromones detected by the opposite type triggers a MAPK cascade. This causes a series of physiological changes where protrusions

in the cells will meet, the cells and some organelles will fuse, and then they will form zygotes (Merlini, Dudin, & Martin, 2013). Any divergence in this pathway may lead to interspecific mating incompatibilities. For example, if two genes from the spindle pole body are incompatible, karyogamy may not be successful, or may form an inviable hybrid unable to undergo mitosis.

Quantitative mating assays between *S. cerevisiae* and *S. paradoxus* have shown that there is a clear preference for *S. cerevisiae* to mate with itself, rather than with *S. paradoxus;* however, the converse is not true and *S. paradoxus* also slightly prefers *S. cerevisiae* (Maclean & Greig, 2008). Hence mating preference alone does not cause complete reproductive isolation, though it may contribute to it.

There are many other means of speciation, and forms of sympatric speciation have been well documented among both animals and plants. One form of sympatric speciation is allochronic speciation, where the timing of essential mating functions becomes so different in a population that two different species arise. Examples include flowering time in plants and diel activity in moths (DEVRIES, AUSTIN, & MARTIN, 2008; Ellis, Weis, & Gaut, 2006; Murphy & Zeyl, 2012). In yeast, *S. cerevisiae* has a shorter germination time than *S. paradoxus*, which diminishes interspecific mating once spores are exposed to rich laboratory media because *S. cerevisiae* is able to mate sooner (Murphy & Zeyl, 2012). Though such conditions do not recapitulate the heterogeneity in the wild, they still represent a partial barrier to gene flow. Other examples of sympatric speciation occur when pheromones diverge in a population (McElfresh & Millar, 2008). The yeast pheromone is relatively simple and huge divergences between pheromone and receptor are tolerated. When pheromones that diverged before the WGD were placed in *S. cerevisiae,* they were able to mate with nearly normal mating efficiency, showing this is not a source of genetic isolation in *Saccharomyces* (Rogers, Denton, McConnell, & Greig,

2015). Hence, among the known methods for sympatric speciation, only allochronic speciation via germination time is a plausible contributor to genetic isolation, although surely this would not cause complete isolation in a heterogeneous wild setting.

The lack of a sharp dividing line between genetic flow is perhaps surprising because *S. cerevisiae* is only thought to outcross once in 50,000 generations (Ruderfer, Pratt, Seidel, & Kruglyak, 2006), allowing ample time for barriers to arrive via divergence. Perhaps all of the relatively small contributions add up to make gene flow so vanishingly rare that two clades in effect have complete reproduction isolation from one another. Nonetheless, no such point has satisfactorily been established, and work continues to find that dividing line.

THE CONTRIBUTIONS OF PROTEIN INTERACTIONS TO HETEROSIS IN HYBRIDS

To what extent do proteins from the different parental species interact in the hybrids? This question has begun to be addressed in interspecific hybrid *Saccharomyces,* and two groups using different methods showed that these diverged proteins do interact with each other. The first of these groups attached two halves of dihydrofolate reductase to proteins from different parents in the nuclear pore complex and RNA polymerase II in *S. kudriavzevii* x *S. cerevisiae* hybrids. This showed that many proteins from the different parental species are in close enough proximity to recapitulate the activity of this enzyme (Leducq et al., 2012). The second study showed that if one component of a hybrid protein complex is tagged, proteins from the other parental species in that complex will be detected after coimmunoprecipitation (CoIP) and mass spectrometry. They found this was true for both *S. cerevisiae* x *S. uvarum* and *S. cerevisiae* x *S. mikitae* hybrids. This group then showed that these heterogeneous complexes can, in some cases, be responsible for

heterosis (Piatkowska, Naseeb, Knight, & Delneri, 2013). Together, these two studies have shown that diverged proteins can interact in hybrids, though to what degree remains an open question.

Interspecific protein interactions are, of course, governed by other principles like molecular forces and protein expression in the hybrids. In interspecific hybrids, protein expression tends to be more additive than in intraspecific hybrids (Blein-Nicolas et al., 2015). This suggests that each species contributes its own way to control protein abundance to the hybrid. Such species-specific control of protein abundance might also contribute to heterosis in a dominant way.

HYBRIDS ARE A USEFUL MODEL TO UNDERSTAND GENE EXPRESSION REGULATION, AND ITS CONTRIBUTIONS TO HYBRID FITNESS

Dominant gene regulation by each parental species also contributes to heterosis in hybrids. Within the hybrids the genome from each parental species is exposed to the same *trans*-environment. Thus comparing expression to purebreds allows one to distinguish between *trans*- and *cis*- modes of gene regulation. In the laboratory setting, *S. cerevisiae* x *S. paradoxus* hybrids have been used to study such expression differences. These hybrids show that *cis*-acting effects tend to be ortholog specific, while *trans*-acting effects are the primary drivers of expression responses to environments in purebreds (Tirosh, Reikhav, Levy, & Barkai, 2009). Expression has also been studied in the hybrid species *S. pastorianus,* and the vast majority of the genes show no difference in expression between orthologs (Horinouchi et al., 2010). However, there are a subset of genes, including sulfur metabolic genes, cellular import genes, and some genes involved in amino acid synthesis, that show an increase for expression of the *S. eubayanus* ortholog. Hence

these studies have shown that differences in expression between orthologs are due to dominant *cis*- effects, they elucidate whether *cis*- or *trans*- effects are acting in the parental species, and show differences in regulation depending on environment.

INTRASPECIFIC HYBRIDS SHARE MANY PROPERTIES WITH INTERSPECIFIC HYBRIDS

Hybrids between different *S. cerevisiae* strains have also been used to study differences of gene expression. These studies utilize the same properties of the intraspecific hybrids as those that use interspecific hybrids and answer many similar questions. For example, these systems allow one to look at divergence in *cis*-regulatory sequences in a controlled *trans*-environment. In particular a BY (closely related to S288C, the standard laboratory strain) x RM11 (a wine strain) hybrid is useful because RM is 99.5% similar to BY, so that it is suitably diverged for many polymorphisms to exist, but close enough that a relatively small number of segregants is enough to statistically unlink virtually every SNP in the genome (Ehrenreich, Gerke, & Kruglyak, 2009). Studies have looked at broad expression differences between BY and RM and their hybrids. In such hybrids, one study found 1,500 alleles that were differentially expressed between the parents (Brem, Yvert, Clinton, & Kruglyak, 2002). Since expression has well-known effects on phenotype, they also investigated how environment affects gene. They found expression of thousands of genes was dependent on strain, environment, or both strain and environment effects (E. N. Smith & Kruglyak, 2008). They also found that *trans*- acting effects tend to be more conserved between environments than *cis*- acting ones (Ehrenreich et al., 2009). These studies highlight the strong impact of gene by environment and gene by genetic background interactions on expression.

Intraspecific hybrids have also been used to understand architecture of quantitative trait loci (QTL). When QTL are mapped to specific genes, reciprocal hemizygosity analysis can be used to determine which allele is contributing to heterosis. Acting in a dominant way, a single allele may be entirely responsible for conferring a phenotype. Using hybrid yeast to demonstrate the importance of reciprocal hemizygosity analyses will have implications for any outbred organism, including humans (Steinmetz et al., 2002).

Hybrids are biologically and economically important, and they provide a wealth of knowledge about basic biological questions. Interspecific hybrids play critical roles in society, and yeast hybrids in particular are essential for the popular alcohol industry. Understanding the biology of these hybrids leads to advances in many fields, including gene expression, the forces that govern protein interactions, and what causes speciation. Nonetheless, much remains to be discovered in these vitally important organisms. Our understanding of the molecular interactions in hybrids remains woefully obscure, and the question about how diverged proteins interact other than a few qualitative observations is largely unknown. Additionally, genome-wide studies, like those done with the *S. cerevisiae* deletion collection, have not been done in hybrids. Finding the genome-wide fitness effect of hemizygous deletions and how these interact with the environment would build off the wealth of knowledge already established in *S. cerevisiae.* Lastly, understanding what could cause complete reproductive isolation with the *Saccharomyces sensu stricto* would answer significant open questions about genetic isolation in this important clade.

CHAPTER 2: *FITNESS EFFECTS OF LOSS OF HETEROZYGOSITY IN HYBRID* SACCHAROMYCES *ARE DEPENDENT ON ENVIRONMENT, GENETIC BACKGROUND, AND SPECIES OF ORIGIN*

Short Title: *Genome-wide fitness consequences of LOH in hybrids*

Author List: Samuel M Lancaster[1], Celia Payen[1], Caitlin F Connelly[1,2], Maitreya J Dunham[1]

Author Affiliations: 1: Department of Genome Sciences, University of Washington, Seattle, WA 2: Current Address, Department of Biochemistry & Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA

Corresponding Author: Maitreya J Dunham

Author contributions:

Caitlin Connelly made the diploid *S. cerevisiae* Tet-off collection

Celia Payen competed the *S. cerevisiae* heterozygous collection *en masse* and sequenced the barcodes.

Sam Lancaster created all the hybrid collections and the diploid *S. cerevisiae* TS collections, did the reciprocal hemizygosity analysis, the individual competitions, all the analyses, and writing.

Maitreya Dunham generated expression and developed the concept of the experiment.

ABSTRACT AND SUMMARY

Abstract: Populations of hybrid species, such as wheat and *Saccharomyces*, are abundant and important. Reminiscent of cancer cell populations, these hybrids have significant genomic rearrangements, including regions of hemizygosity. To understand these rearrangements, we found the genome-wide fitness effects of hemizygosity in hybrid *Saccharomyces* by mating the deletion collection to *Saccharomyces uvarum*, and then competing each strain in three media. We created WT hybrid controls for comparison to the hybrid deletion collection and strains with conditionally knocked out essential genes. We found that the fitness effects of hemizygosity are highly dependent on genetic background, environment, and species of origin of the deleted allele, with significantly less difference in fitness effects within essential genes. Further, we found that hybrids, which display hybrid vigor, have a larger distribution of fitness with more pleiotropy than the purebreds, perhaps explaining why selection often favors them. These findings highlight the enduring importance of experimental biology because predicting phenotype from genotype in hybrid *Saccharomyces* is highly dependent on genotype by environment interactions, and allele by genetic background interactions.

Author Summary: Hybrid yeast are common brewing strains and may offer unique insights into biology and evolution. These hybrids often have chromosomal rearrangements, which we sought to understand on a genome-wide scale. Our work is the first performing a genome-wide scan of single allele deletions in an interspecific hybrid. Such studies are important for understanding copy number variation in diploid organisms with heterozygous genomes. We shed light on these copy number effects in hybrids and why hybrids have such unique karyotypes. Our findings suggest that for every single gene deletion, except for some essential genes, the fitness effects are

highly unpredictable and are dependent on genetic background, environment, and from which parent species the deletion was made. These varied insights may, among other things, explain why hybrid *Saccharomyces* are so often found in human-made environments, where narrow selective pressures select for strong, specific phenotypes.

INTRODUCTION

Hybrid organisms are common in nature, though primarily in plants and fungi, not metazoans. They are also created via artificial selection in agriculture, industry, and the laboratory. In any of these cases hybrids can form a lineage distinct from either parental species in a process known as hybrid speciation (de Barros Lopes, Bellon, Shirley, & Ganter, 2002; Liti & Louis, 2005; Otto & Whitton, 2000). Plant hybridization is common as an estimated 4% of flowering plants and 7% of ferns are hybrids (Otto & Whitton, 2000). Even wheat, a pillar of Western civilization, is a triple hybrid between three grass species (Brenchley et al., 2012). Hybridization is also abundant in budding yeast, both in nature and industrial settings, and many strains show evidence of hybridization and introgression. Simple development and more simple mating systems likely help these yeast tolerate hybridization. The *Saccharomyces sensu stricto* complex is defined as all the species that can successfully hybridize with the budding yeast *Saccharomyces cerevisiae,* and it includes yeasts that are up to 20% diverged (Kellis, Patterson, Endrizzi, Birren, & Lander, 2003). Two common yeasts that originated as hybrids between *S. cerevisiae* and cryotolerant species have even received their own designation as species: the wine species *Sacchaomyces bayanus,* a triple hybrid between *Saccharomyces cerevisiae, Saccharomyces uvarum*, and *Saccharomyces eubayanus;* and *Sacchaormyces pastorianus,* which is primarily a hybrid between *S. cerevisiae* and *S. eubayanus* (Dunn & Sherlock, 2008; Libkind

et al., 2011; Martini & Martini, 1987; Nguyen et al., 2011; Pérez-Través et al., 2014). As in plants, yeast hybrids can occur not just as allodiploids, but also as allotetraploids – a genetic state that may help facilitate forming a stable hybrid lineage (de Barros Lopes et al., 2002; Liti & Louis, 2005; Pfliegler et al., 2012). Yeast hybrids have also been found to shed large portions of their genomes from one or both species (Csoma, Zakany, Capece, Romano, & Sipiczki, 2010; Otto & Whitton, 2000; Peris, Lopes, Belloch, Querol, & Barrio, 2012), they can backcross to introgress diverse genes into one parental species (Dunn & Sherlock, 2008; Libkind et al., 2011; Martini & Martini, 1987; Muller & McCusker, 2009; Naumova et al., 2005; Nguyen et al., 2011; Pérez-Través et al., 2014), and they can mate with other species making triple or quadruple hybrids like with *S. bayanus* (González, Barrio, Gafner, & Querol, 2006; Sipiczki, 2008). Thus, yeast hybrids can have extremely complex genomes.

To study evolutionary forces affecting yeast genomes, experimental evolution has proven effective (Gresham et al., 2008), and hybrid yeast have been experimentally evolved in a variety of media with different selective pressures (Kunicka-Styczyńska & Rajkowska, 2011; Piotrowski et al., 2012). These evolved strains show extensive genome rearrangement that is dependent on the selective pressure used, including loss of heterozygosity (LOH) events that create hemizygous lineages (Kunicka-Styczyńska & Rajkowska, 2011; Piotrowski et al., 2012). Further, these genomic rearrangements are similar to those found in fermentation strains and in some aggressive cancers (Pellman, 2007). However, the factors that drive these genome rearrangements are not well understood. Studying these genome rearrangements, including hemizygosity, in hybrid yeast may yield insights into these phenomena.

Hemizygous loci have been extensively studied in purebred *S. cerevisiae* by using the heterozygous deletion collection that has a single allele deleted for every gene (Deutschbauer &

Davis, 2005; Giaever et al., 2002). Such studies have illustrated that some hemizygous loci cause fitness defects ("haploinsufficiency") or fitness increases ("haploproficiency") in an array of conditions. We expanded these studies to hybrid *Saccharomyces* in order to better understand the effects of hemizygosity in several environments.

To accomplish this we created 8,000 *S. cerevisiae* x *S. uvarum* hybrid yeast strains by mating the nonessential gene deletion collection plus three other collections (one WT control collection and 2 conditional loss of function collections of essential genes, Table 1) of *S. cerevisiae* to WT *S. uvarum*. These collections of hybrid yeast were then assayed in three chemostat media and on solid media to determine the fitness of each strain. These experiments revealed that the fitness effects of deleting a single non-essential allele are dependent on the genetic background, they are unique to the environment with more gene by environment interactions than the purebred, and they depend from which species the original orthologous allele was deleted. For essential genes, however, there was significantly more consistency of genetic effect than between environment and genetic background for non-essential gene deletions. Our findings illustrate the combinatorics of allele by environment and allele by genetic background interactions that affect phenotypes in yeast, and they illustrate the difficulty of performing genotype to phenotype predictions in these important organisms. This study also provides experimental evidence partially explaining genetic heterogeneity in wild and industrial *Saccharomyces* hybrids by showing that particular combinations of alleles and genetic backgrounds generate unique phenotypes suited to particular environments. This use of the strain collections also provides insights into novel phenotypes associated with previously uncharacterized genes, and it demonstrates that hybrids offer a unique fitness landscape with potentially more beneficial mutations, which may contribute to their unique ability to adapt.

RESULTS

We sought to discover the genome-wide fitness effects of hemizygosity in hybrid *Saccharomyces* in three different nutrient-limited chemostat media. To do this we created thousands of hybrid yeast strains with a single allele knocked out or containing a conditional loss of function allele, and compared them to matched purebred strains (Table 1). In parallel, we created control collections of WT *S. cerevisiae* and WT hybrid strains that contain unique DNA barcodes but are otherwise isogenic to allow us to measure technical and biological variation. All strains were then assayed for fitness via competitions in chemostat culture or growth on solid media. Using these approaches, we determined the complement of fitness effects of single allele deletions in a *S. cerevisiae* x *S. uvarum* hybrid, and revealed global properties about the genetics of hybrid yeast.

Table 2.1. Collections used in this study. All collections were created as hybrids and, as controls, diploid *S. cerevisiae.*

Collection	Description
Hemizygous Deletion	The diploid deletion collection, and the haploid deletion collection mated to WT *S. uvarum*. The hybrid hemizygous deletion collection lacks the essential genes.
Barcoded	Haploid barcoded collection mated to both WT *S. uvarum* and WT *S. cerevisiae* (Yan et al., 2008). These are WT control populations for comparison with the deletion collections. Every strain is isogenic except for a ~20mer DNA barcode.
Tet	Haploid tetracycline-off (Tet) collection mated to WT *S. uvarum* and *S. cerevisiae* (Mnaimneh et al., 2004). Essential genes have Tet promoters, and are silenced when exposed to a tetracycline analog.
TS	Haploid temperature sensitive collection mated to WT *S. uvarum* and *S. cerevisiae* (Li et al., 2011). Essential genes are temperature sensitive and are conditionally knocked out at ~37°C.

Fitness effects of LOH in purebred S. cerevisiae *under our experimental conditions*

Genome-scale collections can suffer from background mutations introduced during strain creation and propagation, and methods for quantifying fitness have experimental sources of noise from strain competition, barcode amplification, and sequencing. In order to measure these combined effects in our experimental setup, we generated a distribution of WT *S. cerevisiae* for comparison to the deletion collection. We created 2,000 isogenic yeast that differ only by a unique barcode sequence integrated in the genome, by mating WT BY *S. cerevisiae* to a WT barcoded collection (Yan et al., 2008). To determine the fitness distribution of these strains, we competed them *en masse* in three different nutrient limitations in chemostats, which are continuous culture devices that allow for constant selective pressure. Samples were taken from the chemostats twice daily, cells were pelleted, and genomic DNA isolated. The barcodes from the genomic DNA were then amplified and the change in relative frequency of each barcode over 25 generations was used to determine the fitness of each strain, by determining the slope of the ln(relative population frequency) plotted vs generations (see Materials and Methods). Each experiment was performed in biological replicate (Supplemental Figure 1). We also confirmed that this pooled approach accurately reflects strain fitness by comparing the results to pairwise competitions (Supplemental Figure 4). The fitness of the hybrid barcoded collection strains have a narrow distribution around neutrality, with 98% of the strains in three media falling between a fitness of 0.047 and –0.040 (Fig. 1). These experiments establish the noise one would expect from strain creation, propagation, and our methods for quantifying fitness in a large population of yeast. These cutoffs will denote significance above noise for the analysis of the hemizygous deletion collection.

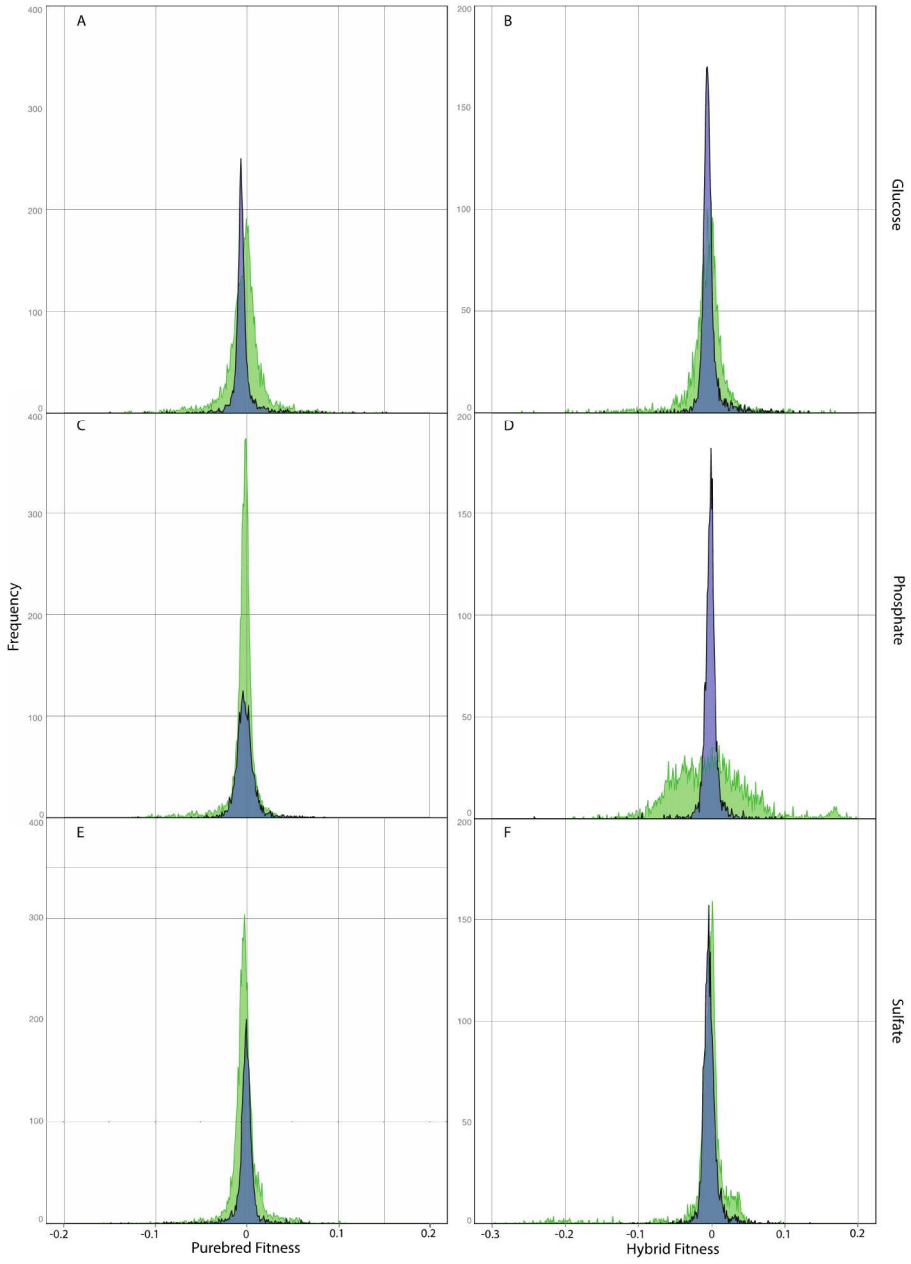

Figure 2.1. Fitness distribution and comparison between WT hybrids and hemizygous hybrids (B, D, F), and comparison of WT purebreds and hemizygous purebreds (A,C,E). Green is the deletion distribution and blue is the WT distribution in (A,B) glucose, (C,D) phosphate and (E,F) sulfate limited media.

With the baseline variation in measurable fitness established, we assayed the hemizygous purebred *S. cerevisiae* deletion collection for fitness the same way as the WT barcoded collection by reanalyzing a dataset published by our group previously (Payen et al 2015, BioRXiv doi: http://dx.doi.org/10.1101/014068). This was done similarly to previous work (Delneri et al., 2008), but with modifications to media and growth rate for comparisons to the hybrids. We found a broader distribution of fitness effects caused by hemizygous gene deletions in *S. cerevisiae* compared to the WT distribution. There were 4,806 strains identified by barcode sequencing in the glucose-limited competition, 4,855 genes in phosphate limitation, and 4,901 genes in sulfate limitation. The null expectation for 1% cutoffs would be 48 extreme fitness strains in each direction in glucose, 49 in phosphate, and 49 in sulfate. However, using the WT cutoffs we observe significantly more strains than the null: 308 haploinsufficient genes and 64 haploproficient genes in glucose limitation ($p<2.2*10^{-16}$, p=0.19); 163 and 5 in phosphate limitation ($p=4.3*10^{-15}$, $p=4.4*10^{-9}$ fewer); and 58 and 113 in sulfate limitation (p=0.44, $p=6*10^{-7}$; statistical tests were Comparison of Two Population Proportions performed in R, with Yates continuity correction). Hence, in the diploid hemizygous deletion collection there are more fitness effects that are more extreme than we would expect than chance (Fig. 1).

Fitness effects of LOH in hybrids

To compare the purebred hemizygous deletion collections with hybrids, we repeated the same set of experiments in the hybrids, including creating a barcoded WT hybrid yeast collection, and 4,000 hybrid hemizygous deletion yeast strains, by mating the deletion collection

to WT *S. uvarum*. Fitness effects were determined in chemostats the same way as the purebred collections.

For the hybrid barcoded collection, like for the purebreds, the vast majority of strains are narrowly distributed around neutral fitness (Fig. 1). For the WT hybrid collection, when the distributions for each nutrient were combined, 98% fell between a fitness of 0.046 and –0.032, which represent the noise we find in the hybrids. This noise is less than we found in the purebreds, though the difference is small. These cutoffs were used to determine outliers in subsequent analyses. For the few WT strains with large fitness effects, there is no fitness correlation between nutrient limitations, indicating that the fitness is medium-specific.

Using the cutoffs determined from the WT collection, we found the strains in the hybrid hemizygous deletion collection with either fitness defects or improvements. In the hybrid collection, there were 3,195 strains identifiable by barcodes in sulfate limitation, 3,179 strains identifiable in phosphate limitation, and 2,955 identifiable in glucose limitation. In the null expectation, based on the cutoffs determined with the barcoded collection, we would expect 32 outliers in each direction in sulfate limited culture, 32 in phosphate limitation, and 30 in glucose limitation. However, in contrast to the purebreds, in the hybrids we observed an excess of outliers in more instances: in sulfate limitation there are 216 haploinsufficient genes and 17 haploproficient genes ($p<2.2 * 10^{-16}$, $p>0.05$); in phosphate limitation there are 919 and 453 ($p<2.2*10^{-16}$, $p<2.2*10^{-16}$); and in glucose limitation there are 308 and 63 ($p<2.2*10^{-16}$, $p=0.0008$; statistical tests were Comparison of Two Population Proportions performed in R, with Yates continuity correction), respectively. The different numbers of outliers between nutrient limited media are reflected in the different shapes of the distributions (Fig. 1). The abundance of

the most haploproficient strains in phosphate limitation by the final time point had necessarily risen to >1.5% of the population, over two orders of magnitude above where they started.

Fitness effects in hybrids are broader than in purebred

To investigate the effect of genetic background on hemizygous deletions, we compared the fitness effects of all hemizygous hybrids and hemizygous purebreds (Fig 1, Fig. 2). The spread of fitnesses in the hybrid background is broader than in the purebreds (Levene test p: < $2.2*10^{-16}$, = $8.6*10^{-6}$, = $1.9 * 10^{-12}$ for phosphate, sulfate, and glucose) suggesting that loss of one allele in hybrids leads to more extreme fitness outcomes, both in the positive fitness and negative fitness directions, than in purebred *S. cerevisiae*.

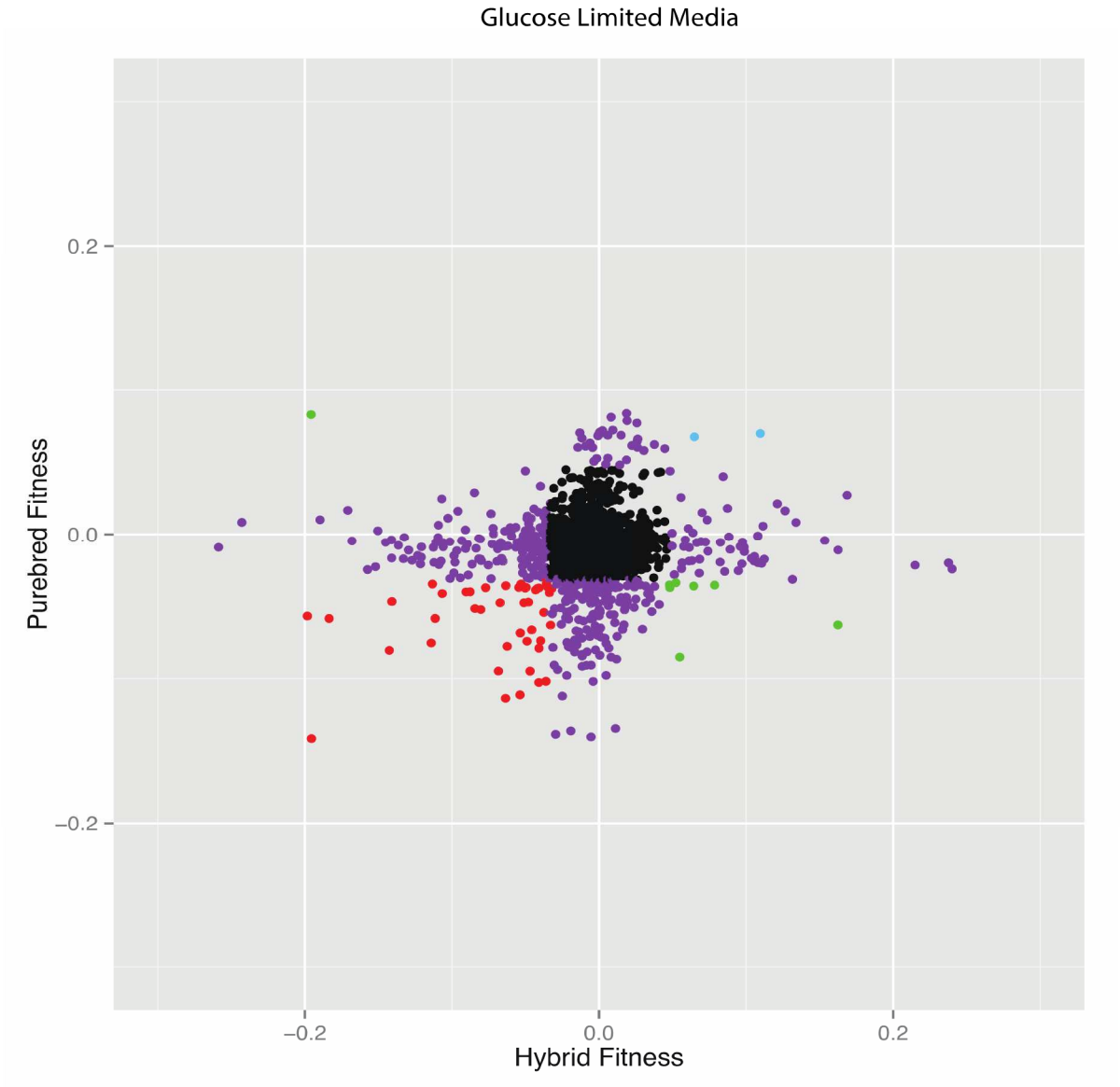

Figure 2.2. Scatter plot of hybrid and purebred non-essential hemizygotes in glucose limitation. Black strains fall inside the 1% cutoff in both axes, purple strains fall outside the 1% cutoff in just one axis, and the other colors fall outside of the cutoffs in both axes. Other nutrient limitation comparisons in Supplemental Figure 3. Data from Supplemental Table 1. $R^2 = 0.00$

Consistent with this lack of correlation, many of the haploinsufficient alleles in the purebred genetic context are alleviated in the hybrid context. We define this as in increase in fitness of at least 0.04 for a haploinsufficient strain, as determined by the negative cutoff in the purebred barcoded WT. In glucose, phosphate, and sulfate limitations there are 93, 54 and 44 such alleviations of haploinsufficiency, respectively. This represents an alleviation rate of 30% and 34% in phosphate and glucose limited media, respectively, and 76% alleviation in sulfate limited media. Gene ontology enrichments for these categories are different than haploinsufficient outliers and include cytosolic ribosomal subunit and ubiquinone metabolic process in glucose limitation; retrograde transport, endosome to golgi, and large ribosomal subunit for phosphate limitation; and the core mediator complex in sulfate limitation. All enrichments have at least $p<0.05$ with a Bonferroni step down correction.

Some strains did have the same fitness effects across backgrounds. Of the 2,775 deletions whose barcodes were present in all experiments for both hybrids and purebreds, none showed haploinsufficiency across all three conditions for both purebred and hybrid genetic contexts, and just four deletions were haploinsufficient in at least 2/3 conditions and neutral in the third in both hybrids and purebreds: *rtg2*, *fen2*, *ygr115c* and *ygr114c*. *YGR115C* and *YGR114C* are dubious open reading frames that overlap with each other and with the 5' end of *SPT6*. *spt6* is not present in the hybrid experiments because a full deletion is inviable, and therefore, its mutant could not pass through the haploid state; however, it appears that a truncated protein is viable, and that it uniformly decreased fitness. *RGT2* is a sensor of mitochondrial dysfunction. Nuclear-mitochondrial incompatibilities have previously been seen in hybrids (Lee et al., 2008), although that does not explain why it is haploinsufficient in the purebreds. *FEN2* is a repressor of carbon

and nitrogen catabolism (Marcireau, Joets, Pousset, Guilloton, & Karst, 1996), and is a symporter used to import small organic molecules (Stolz & Sauer, 1999).

There was only one deletion that was haploproficient in 2/3 environments in both purebreds and hybrids: *ylr280c,* which is also haploproficient in the early stages of purebred wine fermentation (Novo et al., 2013). This ORF is dubious, and lies in a dense cluster of other dubious and uncharacterized ORFs, which may also be contributing to the phenotype. It overlaps with another dubious ORF, *YLR279W,* and with a putative uncharacterized ORF, *YLR281C*. It sits in the 5' region of *YLR283W*, which encodes a putative protein, and also in the 5' of *YLR278C*, a gene for an uncharacterized zinc cluster protein.

Fitness effects of LOH in essential genes are less variable than in non-essentials

To determine hybrid fitness in essential gene hemizygotes, we used two collections of *S. cerevisiae* with conditional loss of function alleles: the Tet collection, where gene expression is inhibited in the presence of a tetracycline analogue (Mnaimneh et al., 2004), and the TS collection, where raising the temperature to ~37°C impairs function of the protein (Li et al., 2011). These collections were mated to WT *S. uvarum* and WT *S. cerevisiae* to create hybrid and purebred heterozygous strains. Each of the Tet collection plates contains three to four WT strains as controls. We further created 96 WT hybrids, making a total of 138 WT hybrid strains as a baseline for comparison. All these WT hybrids grew the same on solid YPD, indicating a false positive rate of <0.7%. In the Tet collection at 37°C there were 51 haploinsufficient hybrid strains, and at 25°C there were 64 haploinsufficient hybrid strains, out of 892 total possible strains. For these strains, 28 overlapped between these two temperatures, and in these different

conditions, this is more than any pairwise comparison between conditions in the nonessentials ($p<2.2*10^{-16}$ for all three pairwise comparisons; Comparison of Two Population Proportions performed in R, with Yates continuity correction). At these temperatures, for the purebreds there were 51 and 50 haploinsufficient strains, respectively; and of these, 39 and 49 were haploinsufficient in both the purebred and hybrid contexts, leaving 12 and 15 hybrid-specific haploinsufficient strains (Fig. 3, Supplemental Figure 1). In the TS collection, 44 genes in the purebred collection were haploinsufficient and 130 of the genes in the hybrid collection were haploinsufficient, with an overlap of 17 between the two genetic backgrounds. This yields 113 hybrid-specific haploinsufficient strains in the TS collection (Supplemental Table 1). Performing a GO analysis on the combined 128 hybrid-specific haploinsufficient genes from all collections shows that they are enriched for organonitrogen compound biosynthetic process, with a p-value of 0.016 after Bonferroni correction.

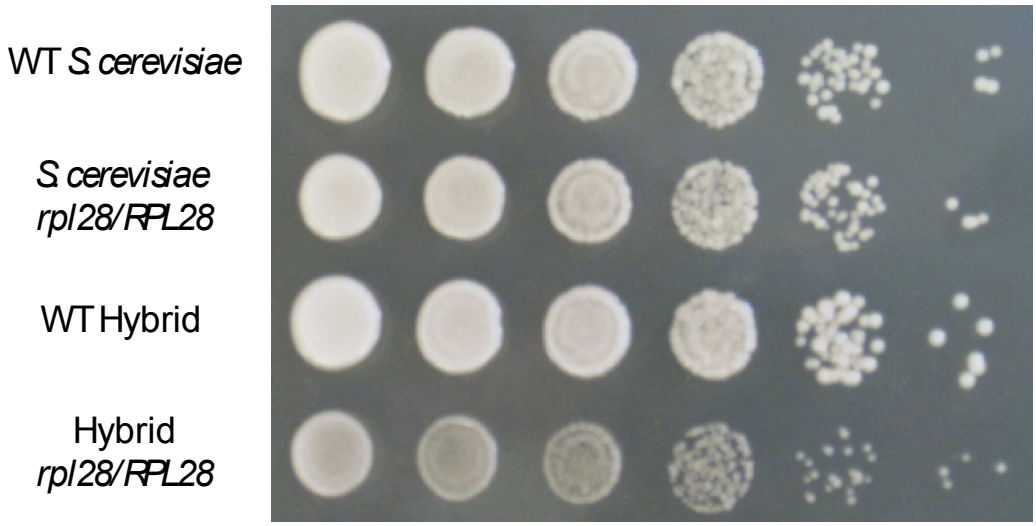

Figure 2.3. Hybrid specific haploinsufficiency on solid YPD. Example strain *rpl28,* in the Tet collection using five 10-fold dilutions. Hybrids have hybrid vigor on solid YPD.

Of the 51 haploinsufficient genes found in the Tet collection grown at the restrictive TS temperature (37°C), only 15 were also present as alleles in the TS collection. Of these 15 present in both collections, 10 were also haploinsufficient in the TS collection, indicating a 2/3 concordance between these collections. These two collections were assayed under different conditions, the alleles were generated in different ways, and many alleles were derived from different backgrounds, showing the consistency of effects. This overlap is more than one would expect by chance ($p=0.002$; Comparison of Two Population Proportions performed in R, with Yates continuity correction) (Supplemental Table 1).

To determine if there were consistent fitness effects between nutrient limitations, we identified all the genes with a significant fitness effect in the purebred *S. cerevisiae* collection in

any of the three nutrient limitations for which data for that deletion was also present in the hybrid experiments. Of the resulting 287 genes, 56 showed the same direction of effects in the hybrid deletion experiment, in either haploinsufficient or haploproficient directions. We compared this number to the essential genes results, and found significantly more overlap in the essential genes between genetic backgrounds ($p=0.01$ for the TS collection, and $p<2.2*10^{-16}$ for the Tet collection; statistical tests were Comparison of Two Population Proportions performed in R, with Yates continuity correction). This indicates that there is significantly less genetic background dependence for the essential genes than in the non-essential genes.

Haploproficiency is not alleviation of dominant interspecific genetic conflict

We hypothesized that one explanation for haploproficiency could be dominant interspecific genetic incompatibility. In this model, deletion of one allele would alleviate the interspecific conflict and confer a fitness benefit to the hybrid. We tested this hypothesis by adding a copy of *S. cerevisiae* haploproficient genes into purebred *S. uvarum.* We transformed *S. uvarum* with 11 plasmids containing haploproficient genes, and then competed these strains individually against *S. uvarum* strains with an empty plasmid. These strains were selections of the haploproficient strains that were also in the Cen MoBY collection (Ho et al., 2009): *PCF11*, *MSH2*, *CDC9*, and *CDC1* in glucose; *TAF11*, *RPL18A*, *MMS22*, and *AMD1* in sulfate; and *TFC1*, *UTP4*, and *CDC1* in phosphate. However, we found no fitness effect in any case (data not shown). These results further confirm that the fitness effects of allele deletions are dependent on genetic background. If haploproficiency is the result of alleviation of a dominant interspecific

genetic incompatibility, there would be a fitness cost when introducing the *S. cerevisiae* allele in the *S. uvarum* background.

Fitness effects of LOH are ortholog-specific

We next hypothesized that the fitness effects we observed might be due to hybrid-specific dosage effects. To test this, we performed a reciprocal hemizygosity analysis to determine if the fitness effects are dependent on the deleted orthologous allele (Steinmetz et al., 2002). If the effects were because of hybrid-specific effects on dosage, we would see the same fitness regardless of which allele was deleted. Hybrid yeast were created with orthologous *S. uvarum* alleles deleted. These were then competed individually against a WT hybrid labeled with GFP. The fitness values of these experiments were completely unassociated with those obtained with the *S. cerevisiae* allele deleted (Table 2). These strains were selected to represent a range of phenotypes from the *S. uvarum* deletions already created by the authors. These results indicate that the effects are likely not due to hybrid-specific dosage (see also gene expression analysis below), and that the effects of fitness are dependent on the deleted ortholog.

Table 2.2. Reciprocal hemizygosity analysis in hybrid *Saccharomyces.*

Nutrient Limitation	Deleted Gene Name	*S. cerevisiae* ORF	Δ *S. cerevisiae* Gene Fitness	Δ *S. uvarum* Gene Fitness
Sulfate	*TMA20*	YER007C-A	-0.18	0
	UTR4	YEL038W	-0.01	-0.01
	PHO13	YDL236W	0.09	-0.19
Glucose	*UTR4*	YEL038W	0.01	-0.01
	PFD1	YJL179W	0.07	-0.02
	YSK1	YJL141C	0.11	-0.01
	RPB4	YJL140W	0.11	0
	SLX5	YDL013W	0.13	-0.01
Phosphate	*UTR4*	YEL038W	0.05	0
	BMH1	YER177W	0.16	0
	YSK1	YJL141C	0.11	0.03
	CPS25	YDR469W	0.17	0.02
	EAF5	YEL018W	0.07	-0.01
	SLX5	YDL013W	0.13	-0.03
	RPB4	YJL140W	0.09	-0.05
	TPK3	YKL166C	-0.15	-0.16

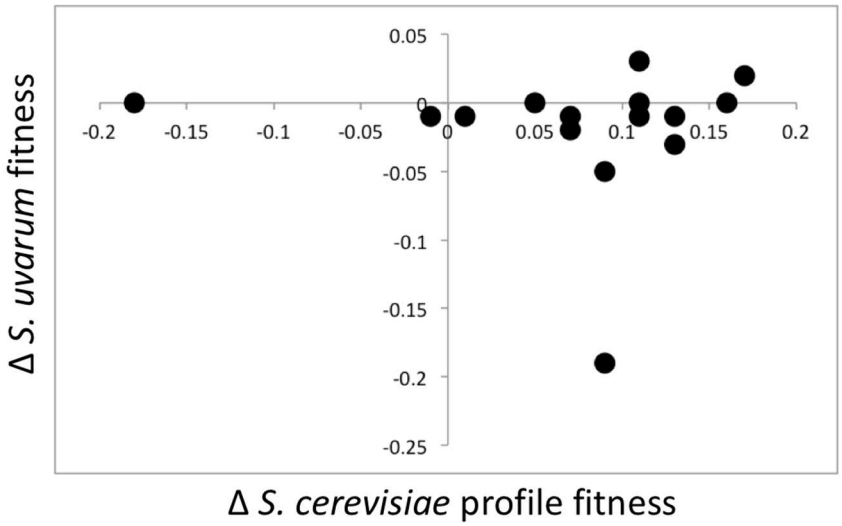

Figure 2.4. Reciprocal hemizygosity analysis correlation between deleted alleles. R^2=0.00.

Heterozygous hybrids have significantly more pleiotropy than purebreds

In addition to background effects, we looked for mutation effects across environmental conditions. When the fitness effects of both hybrid and purebred strains in different nutrient limitations are plotted against one another, we found fitness in one nutrient limitation does not predict fitness in the other (Fig. 4). The genes with antagonistic pleiotropy—low fitness in one medium and high fitness in another—have GO enrichments for gene expression and RNA metabolic process ($p=2.5*10^{-6}$, and $p=4.3*10^{-6}$). We attempted to mitigate the loss of universal haploinsufficient strains by pooling strains after independent growth in 96 well plates, which should have the effect of reducing the relative proportion of strains with antagonistic pleiotropy to as low as possible, yet they were still abundant.

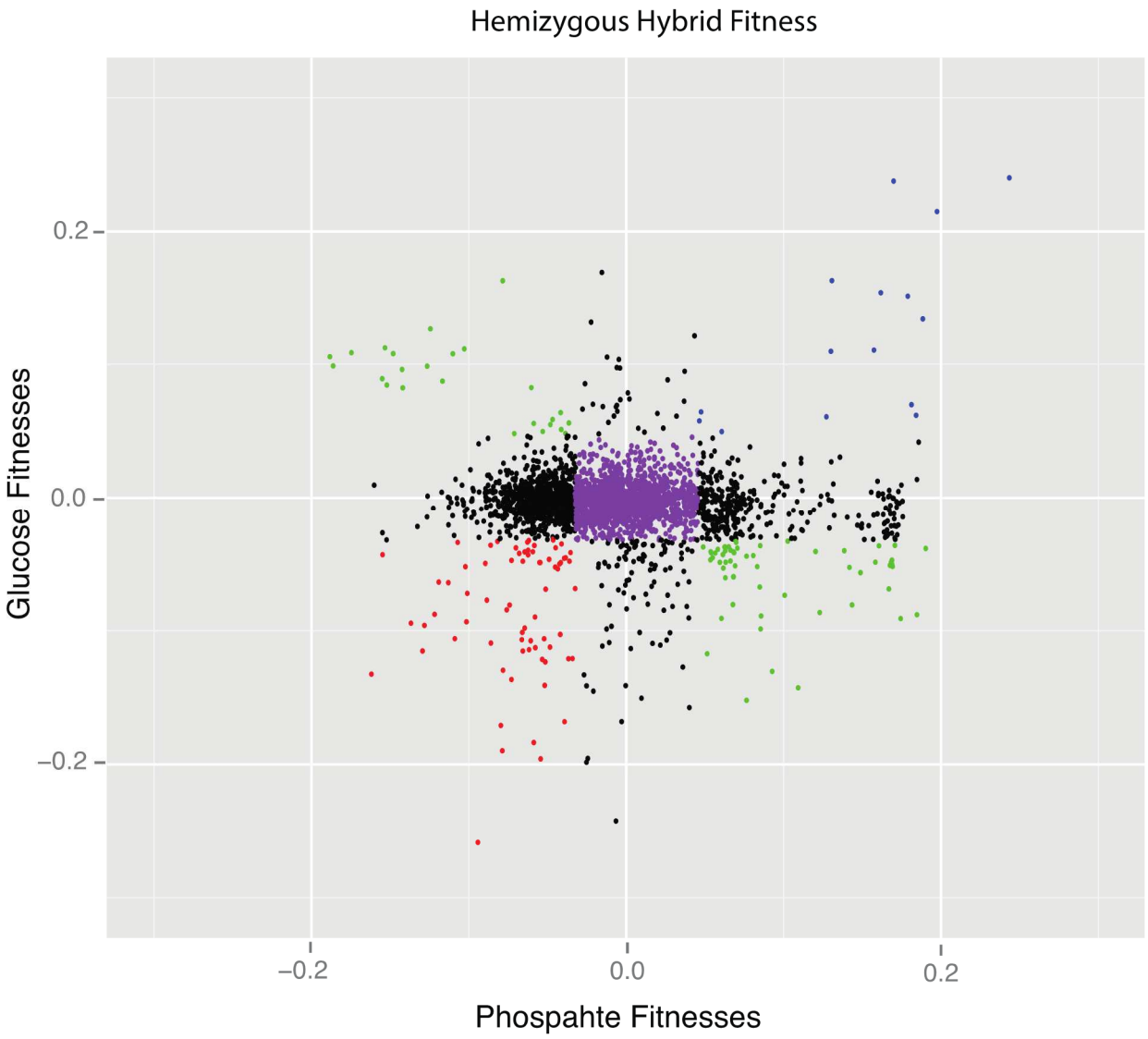

Figure 2.5. Fitnesses of hybrids compared in glucose and phosphate limitation. Strains in black fall inside the 1% cutoff in both axes, purple strains fall outside the 1% cutoff in just one axis, and the other colors fall outside of the cutoffs in both axes. Other comparisons are shown in Supplemental Figure 5. Data from Supplemental Table 1. R^2=0.00

To examine the degree of relative pleiotropy in the purebreds and hybrids, we found the variance among conditions for the 2,775 deletions present in all 6 mass competitions. We did an unpaired T-test between the fitness variances in purebreds and hybrids and found that the hybrid data had significantly more pleiotropy with a p-value of $<2.2*10^{-16}$. Analyzing just glucose and sulfate limited media gave a value of $1.3*10^{-10}$, confirming that even under the conditions with the least variation, there is more relative pleiotropy in the hybrids than the purebreds. We also determined if the broader distribution in hybrids (Fig. 1) accounts for the increased pleiotropy. When performing our pleiotropy test with solely the strains' rank, and not magnitude of fitness effect, there is still increased pleiotropy ($p = 1.6 * 10^{-14}$) using the same methods. Hence, the hybrids have more pleiotropy, regardless of its magnitude.

One contributing factor for the hybrid antagonistic pleiotropy is the absence of sulfate limited haploinsufficient strains in the glucose limited dataset (134/216 strains, $p<2.2*10^{-16}$ based on a random distribution of the missing strains; Comparison of Two Population Proportions performed in R, with Yates continuity correction). This result likely indicates that these deletion strains were so unfit they did not even grow in the chemostats. When only the strains also detected in the glucose limited competition are plotted, the prominent sulfate limited haploinsufficient tail disappears, indicating that these strains have fallen out of the glucose-limited experiment (Supplemental Figure 2).

There were 89 detectable genes that were haploinsufficient in at least two of the media in the hybrid collection, and neutral in the third (Supplemental Table 3). Fewer genes (22) in the hybrid dataset were haploproficient in at least two media and then neutral in the third (Table 3). The fact that there are 92 genes with antagonistic pleiotropy, and 111 with consistent fitness

effects between media in the hybrid dataset is case in point that pleiotropy is common in this non-essential hybrid deletion collection.

Table 2.3. Haploproficiency in hybrid yeast. The last three columns are the fitness profile in the labeled limitation.

Deleted *S. cerevisiae* ORF	Gene Name	Phosphate	Glucose	Sulfate
YBL086C		0.184	0.062	0.001
YBR123C	*TFC1*	0.181	0.07	0.016
YDL013W	*SLX5*	0.131	0.163	-0.007
YDR296W	*MHR1*	0.188	0.134	0.007
YDR325W	*YCG1*	0.037	0.095	0.094
YDR495C	*VPS3*	0.051	-0.012	0.051
YHR025W	*THR1*	0.056	0.002	0.049
YJL140W	*RPB4*	0.13	0.11	0.014
YJL141C	*YAK1*	0.157	0.111	-0.006
YJR064W	*CCT5*	0.128	0.002	0.049
YJR113C	*RSM7*	0.076	-0.011	0.047
YKR079C	*TRZ1*	0.061	0.05	0.023
YLL010C	*PSR1*	0.079	0.038	0.048
YLL027W	*ISA1*	0.047	0.064	-0.011
YLR185W	*RPL37A*	0.127	0.061	-0.012
YLR196W	*PWP1*	0.046	0.058	0.016
YLR280C		0.055	0.045	0.059
YML015C	*TAF11*	0.069	0.011	0.048
YMR142C	*RPL13B*	0.179	0.151	-0.02
YOL090W	*MSH2*	-0.016	0.169	0.057
YOL120C	*RPL18A*	0.085	-0.03	0.05
YOR001W	*RRP6*	0.056	-0.015	0.048

Gene expression differences do not explain hybrid-specific, or environment-specific fitness effects

Large gene expression differences between alleles may affect the importance of dosage (O. C. Martin et al., 2009; Yoon & Brem, 2010). Using a species-specific microarray, we measured gene expression for wild type hybrid strains grown in each of the nutrient limitations. In phosphate limitation, the genes whose deletions were in the top 1% of fitnesses had lower relative expression than the other genes ($p=4*10^{-13}$), and their *S. uvarum* orthologs also had lower expression than the other *S. uvarum* genes ($p=2*10^{-9}$). In phosphate limitation, the *S. cerevisiae* genes with the lowest 1% of fitnesses had no significant difference versus the rest of the genes, but the *S. uvarum* orthologs had significantly higher expression ($p=2*10^{-6}$). No other group of fitness outliers in the purebreds or hybrids was significantly associated with the expression data.

GO enrichments in fitness effects are background dependent

GO enrichments were determined using the 1% fitness outliers in each direction based on the isogenic barcoded collection. Outliers in the purebred data had more enrichments than in the hybrids, as many of those outliers are essential genes, such as components of actin filaments. Both the hybrid and purebred glucose limited haploinsufficient distributions are enriched for gene expression – a rare occurrence where there are consistent effects between the genetic backgrounds. The COPI vesicle protein complex is enriched in the glucose haploinsufficient portion of the purebred distribution, which is consistent with the previous study examining purebreds in chemostats that found the COPI vesicle coat enriched in the haploinsufficient portions of their data (Delneri et al., 2008).

In the phosphate limited haploproficient portion of the purebred data, the most significant enrichments are in chromatin assembly and other functional categories associated with transcription. In the hybrids there are several highly interrelated terms enriched including the GO term macromolecular biosynthetic process. There is also an enrichment in the hybrid haploproficient portion for the SWI/SNF super-type family complex, which regulates growth in response to environmental conditions, among other things. The differences in GO enrichment between hybrids and purebreds highlight the consequences of hybridization on molecular processes. There were no enrichments in the hybrid phosphate limited haploinsufficient portion of the distribution, but this may be due to the high number of strains found outside the isogenic barcoded distribution, diluting out any enrichments.

We found the SWI/SNF super-type family complex across all purebred haploinsufficient enrichments; in contrast, this enrichment is only seen once in a haploproficient portion in the hybrids. Further, in the hybrid sulfate and glucose limited haploinsufficient portions, ribosome biogenesis is enriched, and it is not enriched in any of the purebred distributions. These findings highlight consistency within genetic background, and differences between them.

DISCUSSION

Hybrid yeast are abundant and found in many different types of fermentations. These hybrid yeast, like the hybrids that have been evolved in laboratory conditions, have complex genomic rearrangements, including hemizygosity, reminiscent of cancer cell lines. In an effort to understand what drives hemizygosity in hybrids, we mated the deletion collection and a WT control collection to *S. uvarum*, and then determined the fitness of every strain in several

different media. Comparing fitness between genetic backgrounds, environments, and deleted orthologs provides a picture of the effect of hemizygous mutants in interspecific hybrid yeast.

As controls we competed WT hybrids in three different chemostat media to determine the baseline distribution of fitness for these yeast in our experimental conditions. The distribution was tightly centered around neutrality, and was narrower than for the WT purebreds, but the effect size was small. These experiments were the first to create a baseline distribution fitness in WT hybrid *Saccharomyces.*

We next compared the population of WT hybrid yeast to the hybrid hemizygous deletion collection. The fitness distribution of this hemizygous collection is broader than WT in all conditions. When we compared the distribution of fitness effects with the same experiments performed on purebred *S. cerevisiae,* the hybrids had broader fitness distributions than the purebreds, highlighting that the hybrids more easily move to a larger space of possible fitnesses. When we examine the degree of inter-conditional variation in the collections, we found significantly more in the hybrids than purebreds, and it is independent of the broader distribution of fitnesses. Hence, we have evidence that hybridization combined with hemizygosity further expands the possible fitness landscape *Saccharomyces* can occupy.

These findings of widespread inter-environmental variation in hybrids might be one reason why hybrids are common in manmade environments. In a typical wild setting, cells will cycle between different selective pressures, and antagonistic pleiotropy will prevent any outlier from becoming the most prevalent; however, manmade selective pressures are often narrow, so that the negative pleiotropy may be more limited.

We next sought to interrogate the effects of hemizygosity in essential genes in hybrid yeast using two collections (Tet and TS) that are conditional knockouts. The haploinsufficient

overlap between these two collections is 67% at 37°C. These essential gene collections were created in different ways, and in many cases taken from different strains of *S. cerevisiae*. We found that the overlap of haploinsufficiency between purebred and hybrid backgrounds is significantly greater in essential genes than in the non-essentials. The overlap between environments is also significantly greater in essentials than in non-essentials. Together, these findings highlight the robust effects of gene dosage in essential genes between genetic backgrounds and environments, in relative contrast to the nonessential portion of the genome.

We sought several explanations for why we see the fitness effects of hemizygosity, including changes in expression, dominant interspecific incompatibilities, and hybrid-specific dosage sensitivities – none of which adequately explained the observed results, except for gene expression in phosphate limitation. GO analysis did reveal consistencies within genetic background, indicating that although individual nonessential loci do not have consistent effects, broad functions and pathways might have more consistent effects when disrupted. By and large, we show that hemizygosity in hybrid *Saccharomyces* is highly dependent on environment, genetic background, and deleted ortholog, in such a way that highlights the importance of experiments in predicting phenotype.

The broad nature of this study and its novel use of the strain collection resources in *S. cerevisiae* have led to several findings. We confirmed the previously described finding in purebred yeast that fitness effects of single allele deletions have different fitness effects between different environments, and we have shown that this is also true in hybrid yeast; however, we did find that broad GO enrichments had more consistencies. We highlight the importance of reciprocal hemizygosity analysis when understanding the genetic contribution of a locus to a phenotype in heterozygous yeast. We show that hybridization potentially creates a broader

fitness landscape than can be accomplished in purebreds. We confirmed that in purebreds, there is no association between expression and fitness effects; however, in the hybrids we found evidence of some associations between expression and fitness in phosphate limited conditions. Lastly, we show that essential genes, representing the core cellular processes, show significantly more phenotypic consistency between genotype and environment than deletion of non-essential genes.

METHODS

Strains and collections

All *S. uvarum* strains are derived from the type strain CBS 7001, sometimes identified as *S. bayanus* var *uvarum*. Hybrid collections were made by spreading 200 µl mid log *S. uvarum lys2 MAT*α on solid YPD omni plates then spotting the haploid *MATa* collections (Tet, TS, deletion, or WT barcoded), using a pinner with 96 arrayed pins. After overnight growth, colonies were transferred to selective solid minimal media to ensure only hybrid growth. These plates were then transferred to liquid selective media and stored at -80°C. The collections were saved as 96 well plates, and duplicates of the deletion and barcoded collections were saved as pools. *S. uvarum* deletion strains were provided by the Rine lab at UC Berkeley.

Determining Fitness of Essential Genes

The haploid TS collection was first assayed for the effectiveness of the inducible phenotype in our laboratory conditions by spotting them on YPD and then growing them at

37°C. To assay for haploinsufficiency, we performed four 1:10 serial dilutions of each well of the heterozygous diploids. For the Tet collection, these four serial dilutions were plated on solid YPD + 40 μg/ml doxycycline. The TS and Tet collections were plated twice on YPD. Both collections were grown for 24 hours at 25°C and 37°C, respectively (the Tet collection was grown at 37°C for comparison to the TS collection). For the Tet collection, haploinsufficiency was determined by comparing the WT wells to deletion strain wells, and in the TS collection, by comparing between restrictive and permissive temperatures for the same allele. Hybrid and diploid purebred collections were assayed in the same way to determine hybrid-specific haploinsufficiency.

Fitness Assays of Non-Essential Genes

The barcoded and deletion collections, which were created for both the purebreds and hybrids, were assayed as pools in duplicates in three different chemostat media – glucose limited, phosphate limited, and sulfate limited. For comparison to hybrids, experiments with purebreds were taken from Payen et al. (2015 BioRXiv doi: http://dx.doi.org/10.1101/014068), where the protocol is described in detail. Pools were inoculated into 240ml chemostats and grown for 24 hours, when peristaltic pumps were turned on at a dilution rate of ~0.17 volumes per hour. Samples were taken from chemostats twice a day. From these samples, the unique DNA barcodes from each collection were PCR amplified, with each time point having a unique Illumina adapter incorporated during PCR amplification. The barcodes were then sequenced on an Illumina Genome Analyzer IIx. The frequency of the barcodes was used to calculate the fitness of each strain by determining the natural log in the change of proportional barcode

frequency over 25 generations. We required a minimum of 100 barcodes per strain to be identified. For the reciprocal hemizygosity analysis, the 100 barcode limit was reduced to 44 for comparison to the mass competitions because manual curating of these ensured no false positives. Many sequences for the WT barcoded collection DNA barcodes were only determinable through examination of over represented sequences in sequencing data, and these were used for analysis.

Individual competitions were done in the respective media in 20mL chemostats and competed against a single WT clone with a GFP label. Fitness was determined by regressing the slope of generations versus the ln(dark cells/GFP cells).

Gene Expression Analysis

Gene expression data were collected as previously described (O. C. Martin et al., 2009), for every allele in the purebreds and hybrids in all three chemostat nutrient limitations.

GO Enrichments in the Dataset

GO enrichments were determined using the ClueGO application in Cytoscape (Bindea et al., 2009), and using the total strains identified in our experiments as the background population. Outliers were determined using a 1% cutoff in each direction based on the WT barcoded collection. All ontologies were corrected for multiple comparisons with a Bonferroni step down analysis.

SUPPLEMENTAL

.

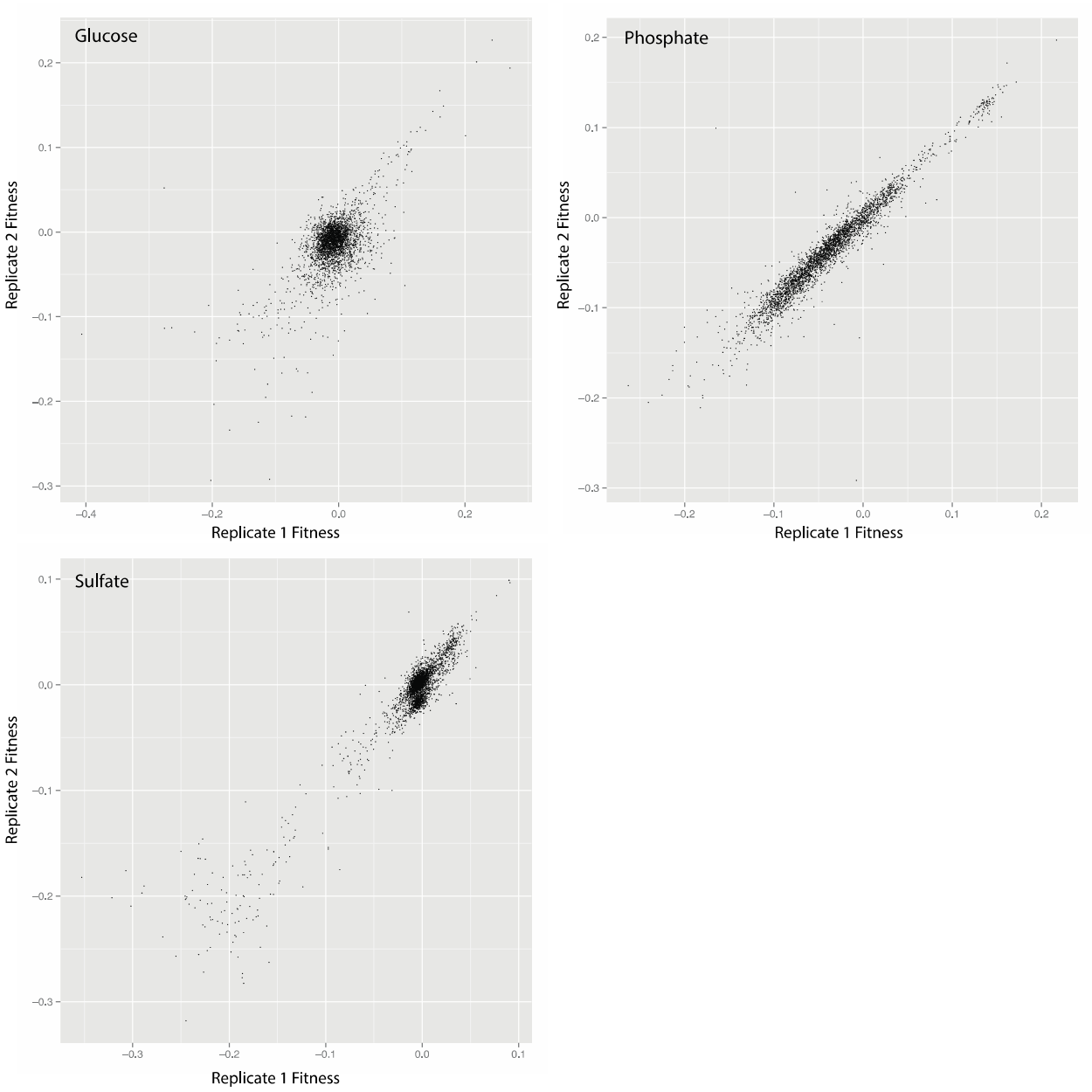

Figure 2.6. Plot of *en masse* hybrid hemizygote replicates in A) phosphate, B) sulfate and C) glucose limitations.

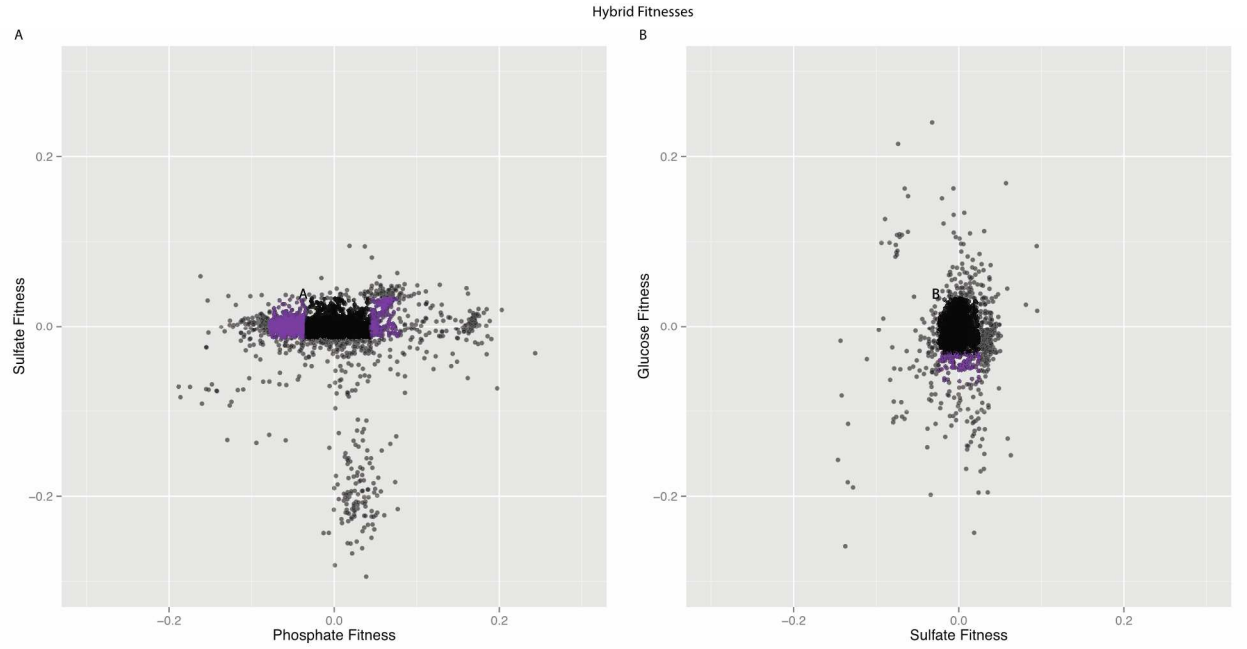

Figure 2.7. Plot of hybrid hemizygous fitnesses between A) sulfate and phosphate limitations and B) glucose and sulfate limitations. Black strains fall inside the 1% cutoff in both axes, purple strains fall outside the 1% cutoff in just one axis, and the other colors fall outside of the cutoffs in both axes. Both $R^2 = 0.00$

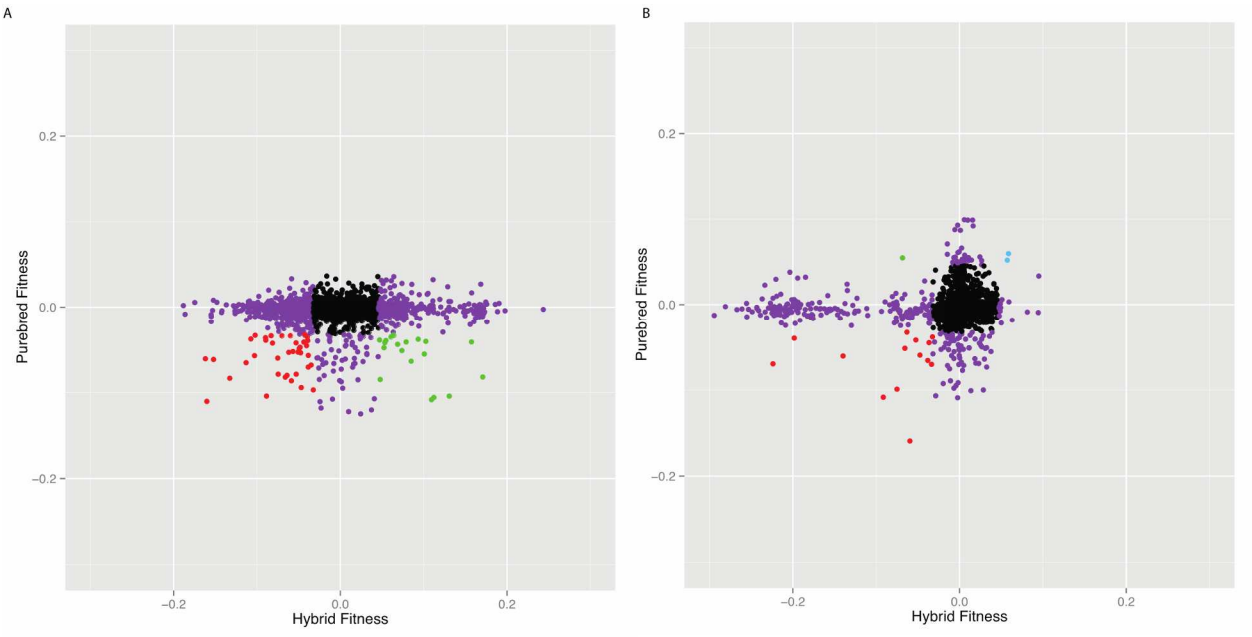

Figure 2.8. Plot of hybrid and purebred fitness in A) phosphate and B) sulfate limitations. Black strains fall inside the 1% cutoff in both axes, purple strains fall outside the 1% cutoff in just one axis, and the other colors fall outside of the cutoffs in both axes. Both $R^2 = 0.00$.

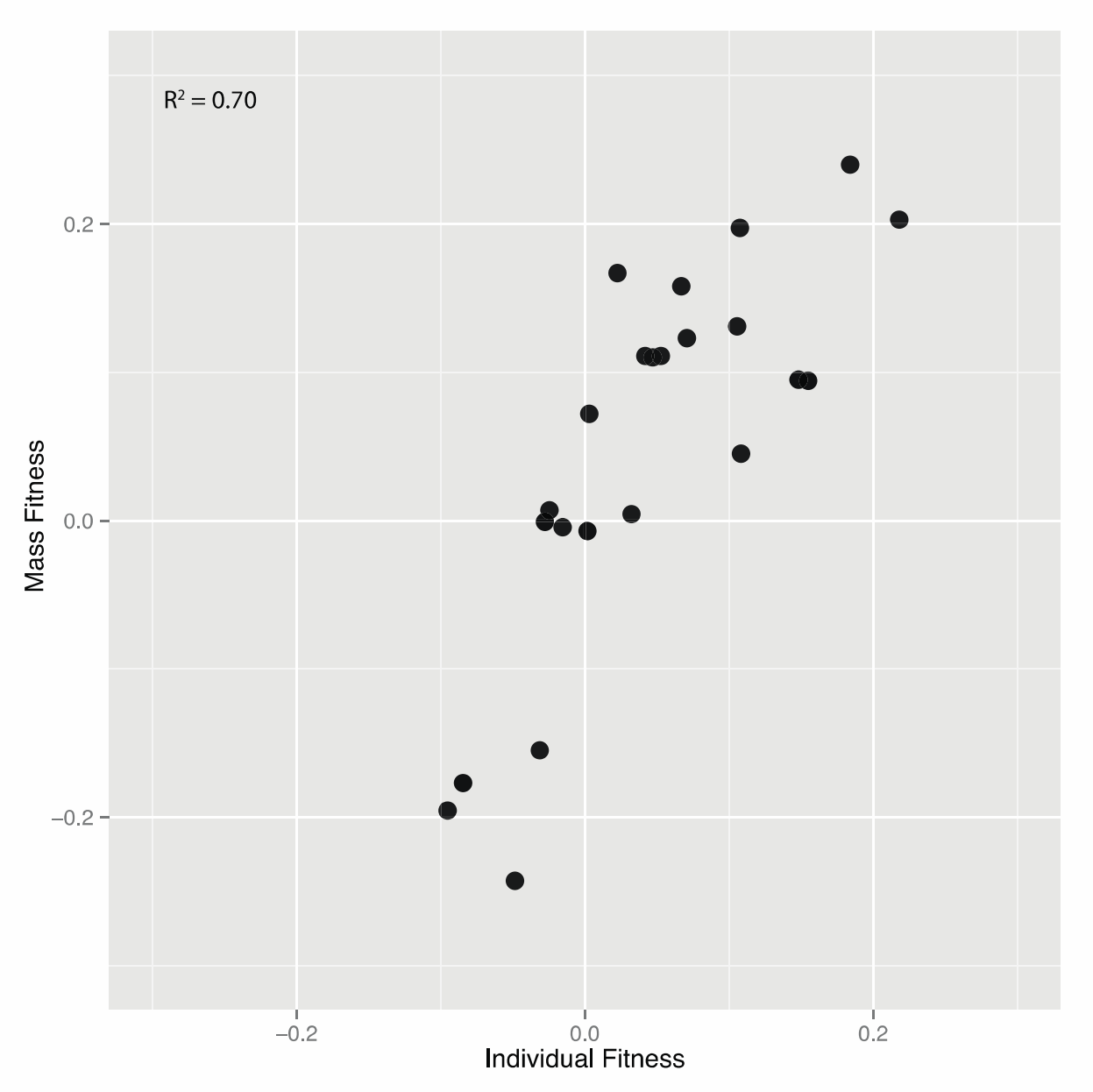

Figure 2.9. Plot of *en masse* barseq fitnesses and individual competitions between mutant strains and GFP-labeled hybrids.

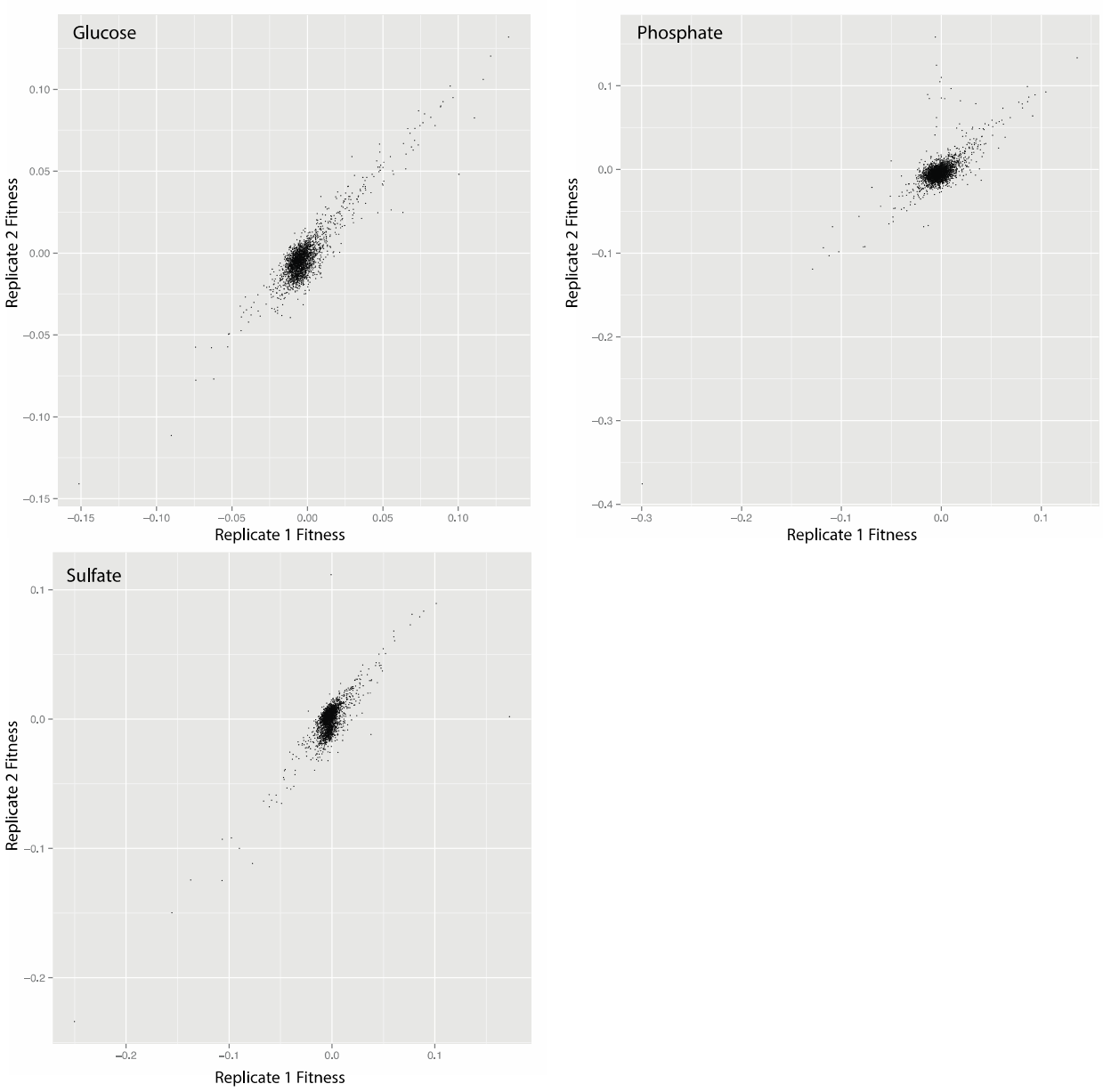

Figure 2.10. Plot of barcoded competition replicates in A) phosphate, B) sulfate and C) glucose limitations.

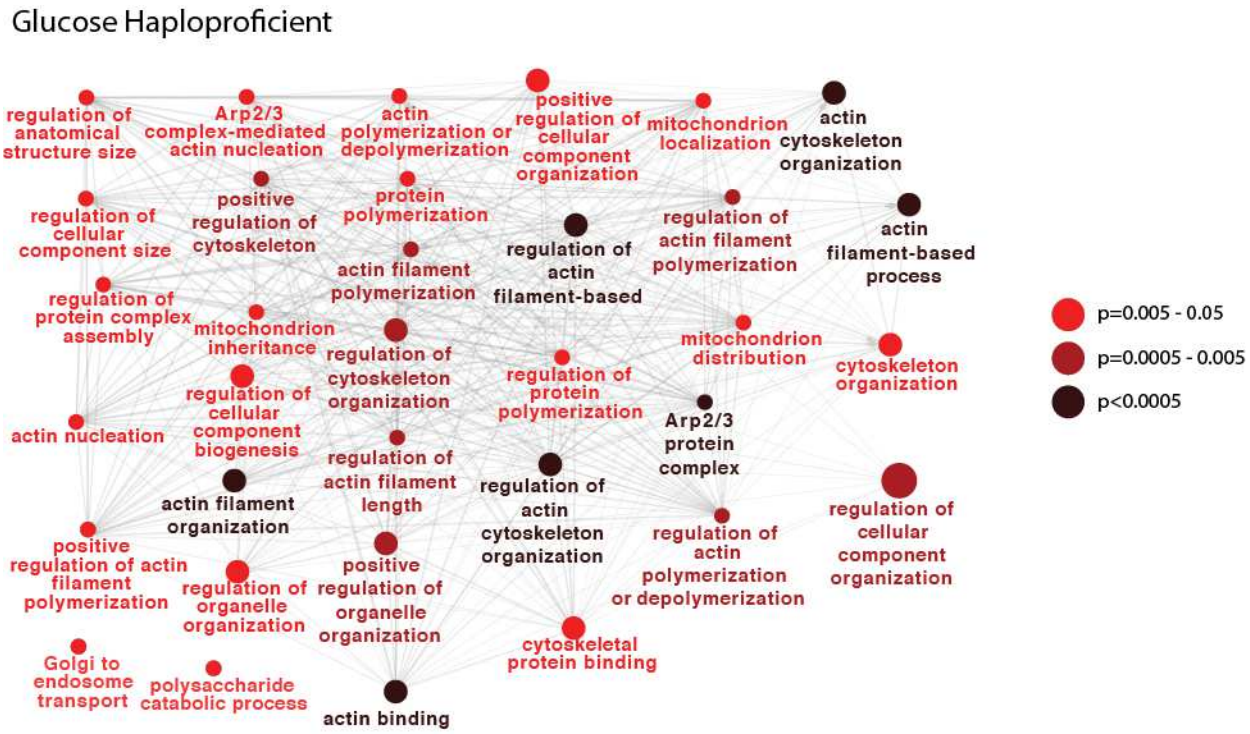

Figure 2.11. GO enrichments in the haploproficient portion of the glucose limited mass competition in purebreds. Edges represent nodes with overlapping genes.

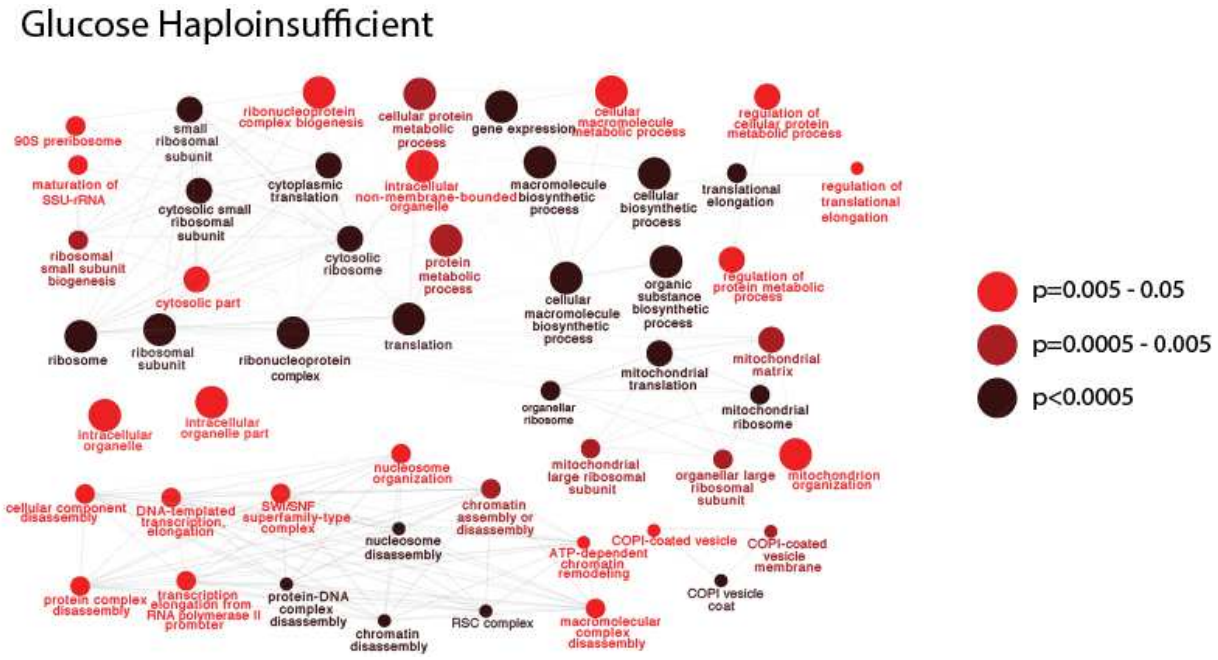

Figure 2.12. GO enrichments in the haploinsufficient portion of the glucose limited mass competition in purebreds. Edges represent nodes with overlapping genes.

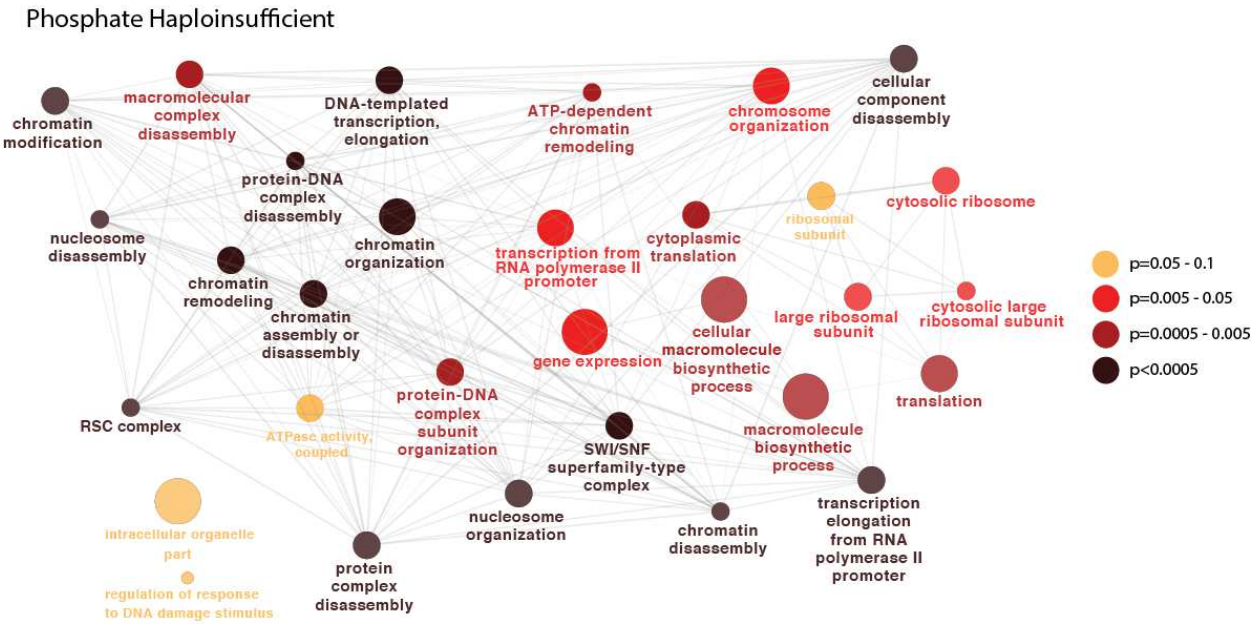

Figure 2.13. GO enrichments in the haploinsufficient portion of the phosphate limited mass competition in purebreds. Edges represent nodes with overlapping genes.

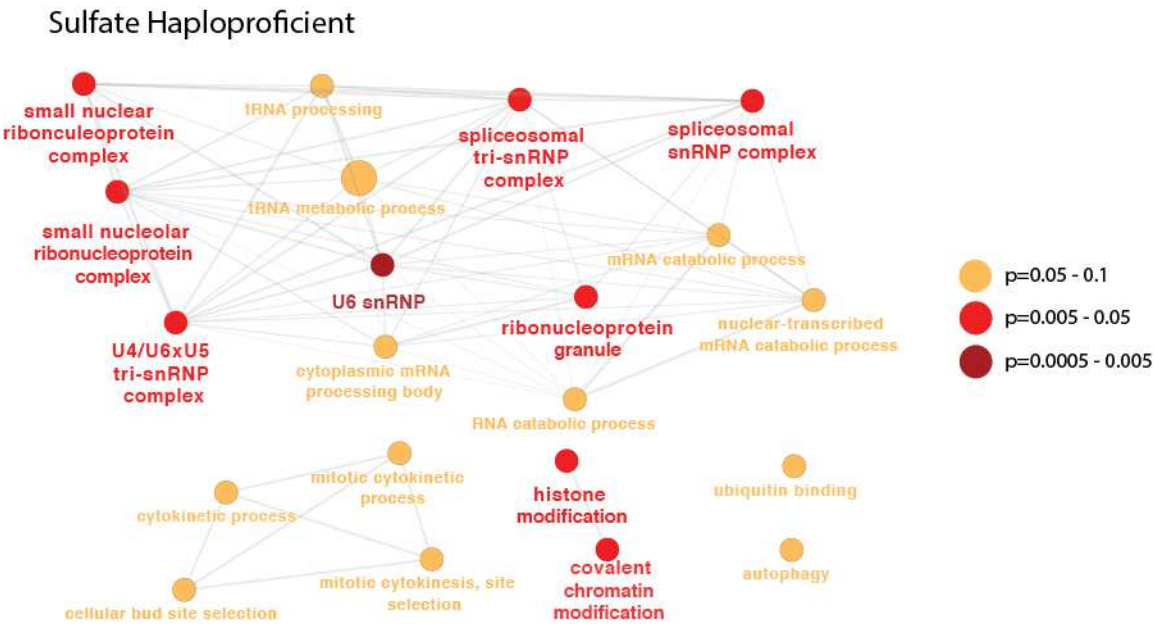

Figure 2.14. GO enrichments in the haploproficient portion of the sulfate limited mass competition in purebreds. Edges represent nodes with overlapping genes.

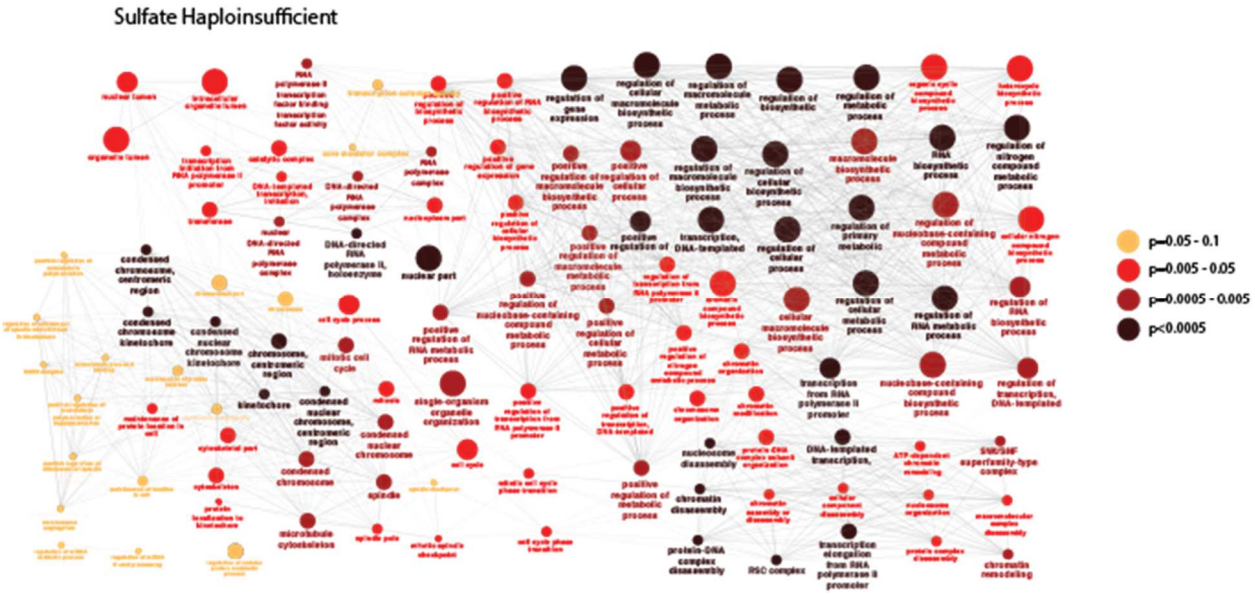

Figure 2.15. GO enrichments in the haploinsufficient portion of the sulfate limited mass competition in purebreds. Edges represent nodes with overlapping genes.

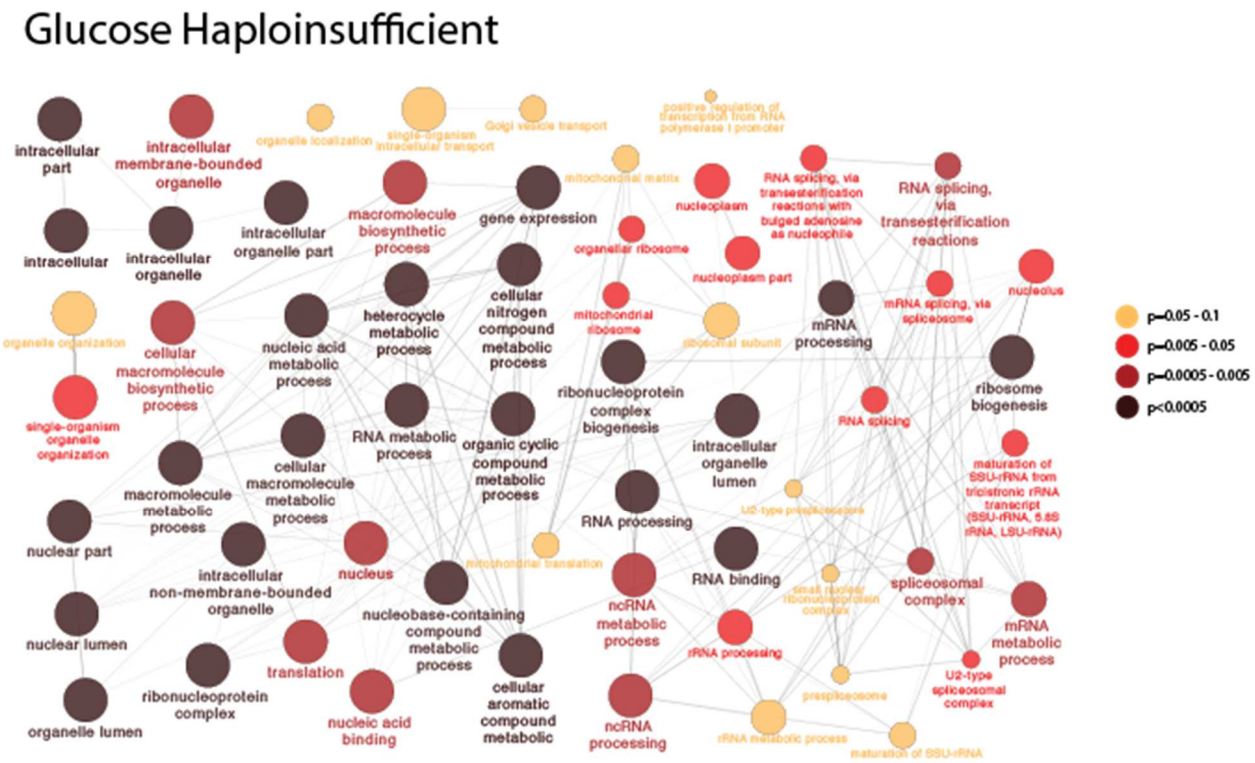

Figure 2.16. GO enrichments in the haploinsufficient portion of the glucose limited mass competition in hybrids. Edges represent nodes with overlapping genes.

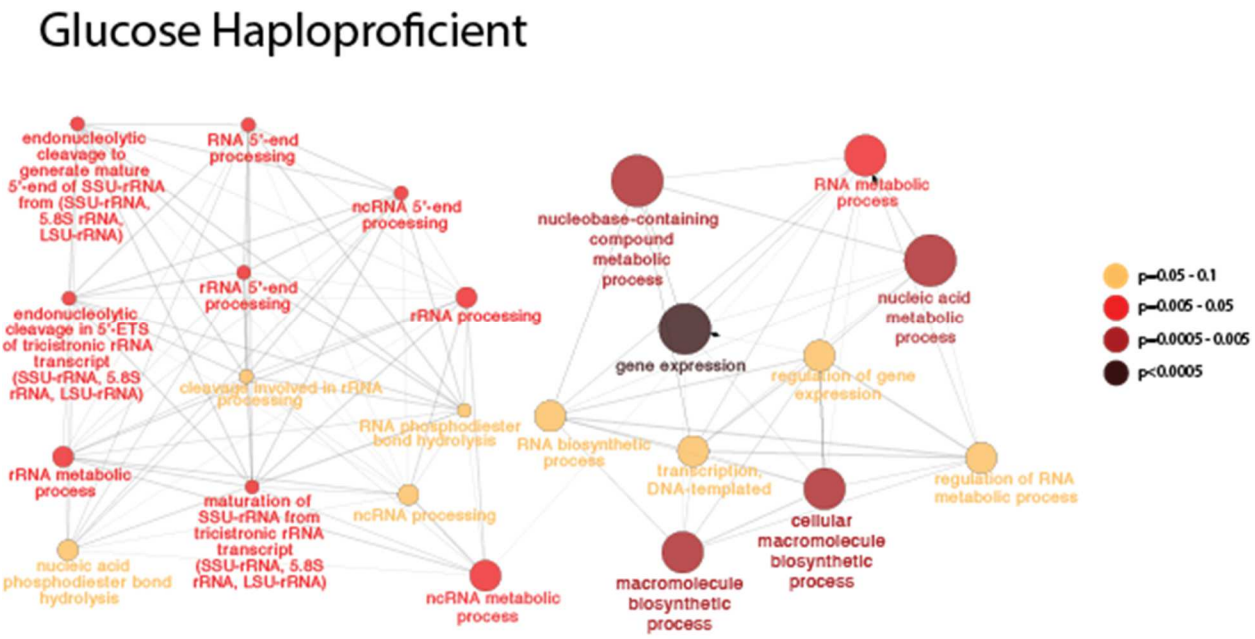

Figure 2.17. GO enrichments in the haploproficient portion of the glucose limited mass competition in hybrids. Edges represent nodes with overlapping genes.

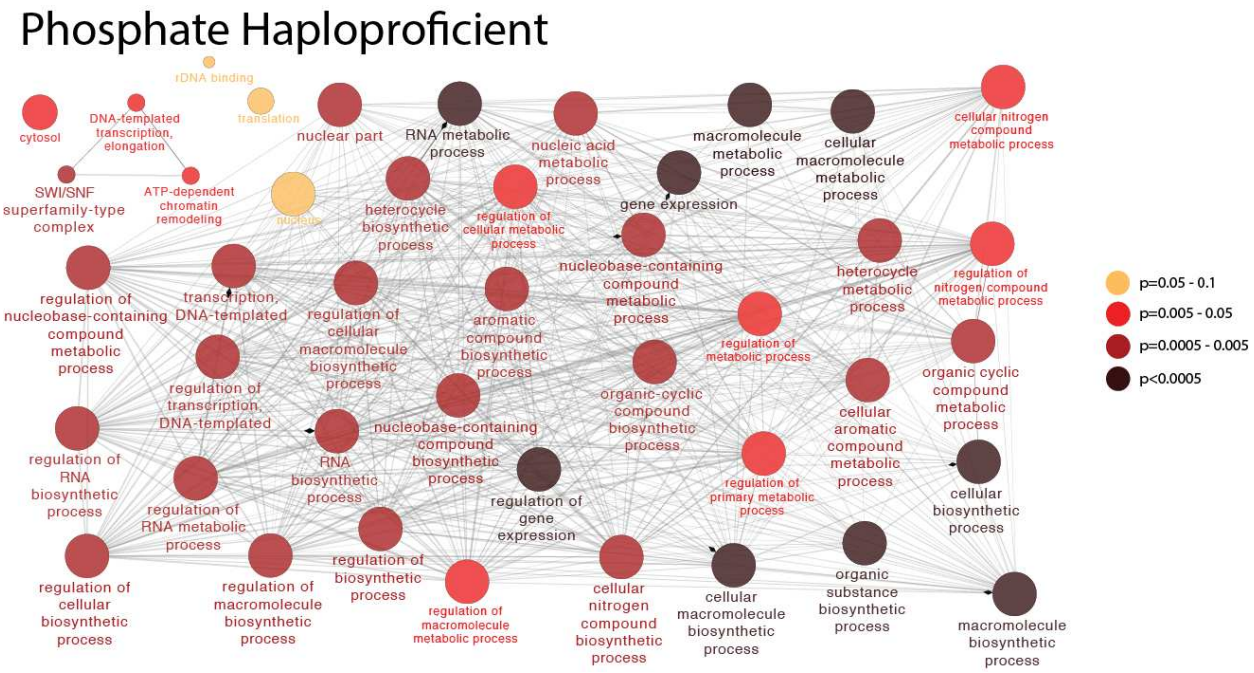

Figure 2.18. GO enrichments in the haploproficient portion of the phosphate limited mass competition in hybrids. Edges represent nodes with overlapping genes.

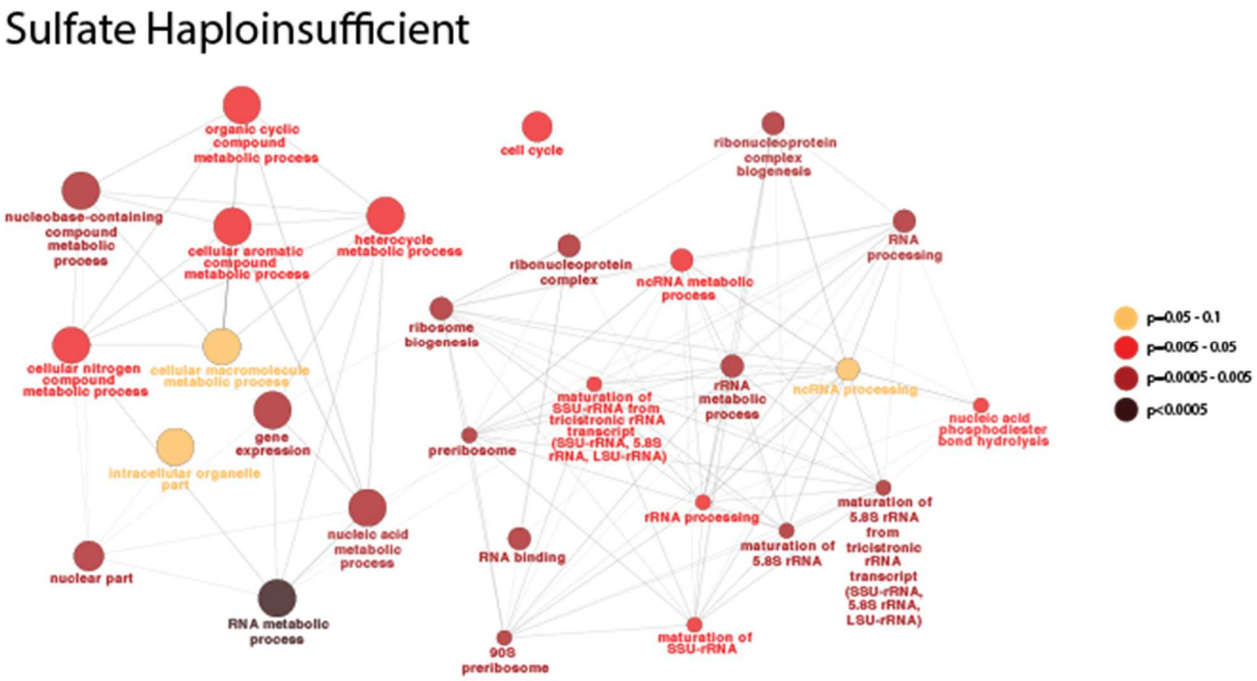

Figure 2.19. GO enrichments in the haploinsufficient portion of the sulfate limited mass competition in hybrids. Edges represent nodes with overlapping genes.

Table 2.4. Haploinsufficient genes in essential gene collections.

TS Hybrid	TS Purebred	Tet Purebred 37C	Tet Hybrid 37C	Tet Purebred Room Temperature	Tet Hybrid Room Temperature
YLR293C	YLR293C	YAL043C	YOL120C	YAL033W	YOL120C
YBR167C	YOR070C	YBL030C	YKR037C	YBR079C	YKR037C
YDL077C	YER133W	YBL084C	YLL050C	YER003C	YLL050C
YCR042C	YOR074C	YBR136W	YPR183W	YER006W	YPR183W
YER148W	YML098W	YBR140C	YPL063W	YER165W	YPL063W
YIL068C	YDR463W	YBR167C	YMR270C	YKL052C	YPL063W
YNR017W	YLR029C	YER006W	YFL029C	YKR083C	YMR270C
YPL055C	YFL039C	YER168C	YMR001C	YFR031C	YFL029C
YNL282W	YML031W	YIL150C	YOR168W	YKL006C-A	YMR001C
YLR071C	YDR398W	YIR010W	YLR272C	YDL047W	YDR016C
YOL069W	YDL014W	YKL009W	YDR407C	YDL105W	YLL003W
YOR294W	YLR321C	YKL052C	YNL221C	YDR016C	YMR005W
YFL034C-B	YAL041W	YKL203C	YNL310C	YDR060W	YLR198C
YLR212C	YJR065C	YKR083C	YNL149C	YDR292C	YDR407C
YLL004W	YDR208W	YFL045C	YLR223C	YDR407C	YLR223C
YBR087W	YNL061W	YFL034C-B	YBR140C	YDR531W	YNL149C
YKR086W	YFR037C	YFR005C	YPL126W	YGL098W	YNL310C
YKL088W	YHR058C	YKL006C-A	YKL009W	YGL103W	YPR082C
YLR186W	YAL003W	YDL016C	YER168C	YGL108C	YPL126W
YOR174W	YJL019W	YDL105W	YHR072W-A	YGR158C	YGR158C
YCL059C	YKL024C	YDR016C	YBL030C	YGR173W	YPR016C
YGR094W	YFL022C	YDR168W	YFL045C	YJR002W	YAL033W
YAL013W	YFL034C-B	YDR356W	YKL006C-A	YJR007W	YHR072W-A
YGL093W	YDL033C	YDR407C	YLR316C	YJR123W	YDL105W
YDL207W	YPR103W	YGL098W	YML126C	YLR005W	YKL006C-A
YKL203C	YGL130W	YGL103W	YMR309C	YLR076C	YLR316C
YGR101W	YFR028C	YGL106W	YGL103W	YLR106C	YML126C
YLR310C	YOR174W	YGR173W	YDL105W	YLR140W	YMR203W
YMR307W	YBR236C	YLR029C	YPL228W	YLR198C	YDR292C
YBL034C	YNL147W	YLR106C	YNL313C	YLR223C	YFR031C
YNL102W	YER093C	YLR140W	YPL231W	YLR316C	YER165W
YOL077C	YGR116W	YLR223C	YOR372C	YLR323C	YMR146C
YGL055W	YDL017W	YLR272C	YGL068W	YML126C	YMR309C
YOR074C	YIL022W	YLR316C	YPL076W	YMR146C	YER006W
YIL150C	YGL093W	YLR323C	YFL035C	YMR203W	YJR002W
YLL036C	YLR378C	YLR397C	YLR323C	YNL149C	YGR173W

YIL062C	YNR010W	YML126C	YGL106W	YNL150W	YBR079C
YIL021W	YDL207W	YMR203W	YLR140W	YNL272C	YDR531W
YKR092C	YGL097W	YNL149C	YIR010W	YNL310C	YJR123W
YOR259C	YHR172W	YNL221C	YDR356W	YOR102W	YKR083C
YDL102W	YFL017C	YNL310C	YAL043C	YOR335C	YJR007W
YML069W	YJL085W	YNL313C	YDR168W	YOR372C	YDR060W
YPR033C	YBL034C	YOR116C	YBL084C	YPL010W	YLR005W
YJR139C	YJR074W	YOR168W	YOR116C	YPL076W	YGL108C
BLANK		YOR259C	YKL052C	YPL126W	YNL272C
YDR087C		YOR372C	YLR106C	YPR016C	YPL010W
YHR085W		YPL076W	YOR259C	YPR082C	YOR335C
YIR010W		YPL126W	YMR203W	YPR107C	YPR169W
YOR244W		YPL228W	YDR016C	YPR169W	YNL272C
YBL084C		YPL231W	YLR397C	YHR072W-A	YOR004W
YKL022C		YHR072W-A	YBR136W		YOR372C
YMR235C			YBR167C		YGL068W
YDR126W			YDL016C		YPL076W
YEL034W			YER006W		YER003C
YNL181W			YFR005C		YPR107C
YOL102C			YGR173W		YLR323C
YGL011C			YKR083C		YLR106C
YDL017W			YKL203C		YKL052C
YLL031C			YIL150C		YLR101C
YBR011C			YLR029C		YOR102W
YDL111C					YGL103W
YGL048C					YNL150W
YNL222W					YLR140W
YIL118W					YLR076C
YDR460W					
YFL017C					
YOR236W					
YOR335C					
YMR224C					
YLR029C					
YFL039C					
YNL287W					
YJL034W					
YNR053C					
YPL255W					
YOL001W					

YLR317W						
YPL228W						
YNR003C						
YMR271C						
YMR296C						
YBR221C						
YJL097W						
YPR103W						
YGL116W						
YFR028C						
YDL028C						
YHR012W						
YNL118C						
YLR378C						
YDL103C						
YBR152W						
YLR215C						
YPR018W						
YLR314C						
YOR048C						
YER018C						
YML001W						
YGL233W						
YKL189W						
YFL045C						
YDL030W						
YDL143W						
YDR246W						
YMR277W						
YIL046W						
YPL101W						
YFL022C						
YOR020C						
YOR254C						
YBR164C						
YDL033C						
YHR034C						
YLL035W						
YER116C						
YKL089W						

YDR356W					
YBR131W					
YDR167W					
YDR328C					
YBR079C					
YOR014W					
YOL038W					
YGR274C					
YGR103W					
YPR016C					
YBR049C					
YDR044W					
YJR074W					
YBR154C					

Table 2.5. Haploinsufficient hybrid strains in two out of three chemostat media.

ORF	Phosphate Limited	Sulfate Limited	Glucose Limited
YGL252C	-0.160	0.010	-0.091
YER082C	-0.155	-0.043	-0.025
YKL045W	-0.137	-0.094	-0.001
YGR065C	-0.129	-0.115	-0.134
YKL012W	-0.128	-0.096	0.036
YHR060W	-0.122	-0.088	0.002
YMR193W	-0.119	-0.064	0.026
YOR140W	-0.113	-0.064	-0.018
YDR207C	-0.109	-0.106	-0.011
YOR354C	-0.107	-0.034	-0.029
YAL043C	-0.102	-0.052	-0.023
YKL052C	-0.102	-0.093	0.036
YKL179C	-0.101	-0.072	0.005
YER022W	-0.095	-0.019	-0.036
YGR013W	-0.094	-0.259	-0.137
YCR003W	-0.090	-0.050	-0.040
YKL006W	-0.089	-0.077	-0.011
YGL244W	-0.086	-0.036	0.013

YOR329C	-0.086	-0.109	-0.066
YDR043C	-0.082	-0.033	-0.001
YKL049C	-0.080	-0.171	0.026
YGR103W	-0.079	-0.190	-0.128
YBR055C	-0.078	-0.130	0.017
YHL015W	-0.076	-0.084	-0.006
YNR051C	-0.074	-0.081	-0.013
YBL004W	-0.073	-0.136	0.015
YLR391W	-0.073	-0.047	0.008
YCR100C	-0.070	-0.038	-0.014
YDL005C	-0.068	-0.042	-0.011
YBR002C	-0.066	-0.107	-0.069
YKL180W	-0.066	-0.101	0.031
YMR057C	-0.066	-0.048	-0.055
YHR065C	-0.066	-0.115	-0.028
YBR079C	-0.065	-0.098	0.011
YHR062C	-0.064	-0.041	-0.016
YHR042W	-0.063	-0.034	0.005
YNL296W	-0.062	-0.043	-0.006
YKL112W	-0.062	-0.040	0.026
YLR278C	-0.062	-0.032	-0.002
YKR002W	-0.062	-0.114	0.017
YBL077W	-0.061	-0.107	0.015
YDL006W	-0.059	-0.041	-0.023
YGR115C	-0.059	-0.184	-0.134
YAL025C	-0.058	-0.036	-0.032
YGL020C	-0.058	-0.090	-0.070
YKR062W	-0.058	-0.113	0.018
YDR025W	-0.055	-0.049	0.011
YNL055C	-0.055	-0.049	-0.006
YKL013C	-0.054	-0.196	0.024
YKR082W	-0.053	-0.121	0.032
YDL178W	-0.052	-0.106	-0.007
YBL026W	-0.052	-0.141	0.010
YBL073W	-0.051	-0.123	0.019
YBL038W	-0.051	-0.069	0.012
YAR007C	-0.049	-0.047	-0.024
YBR117C	-0.048	-0.112	0.010
YMR173W	-0.046	-0.032	-0.006

YJL061W	-0.045	-0.052	0.028
YDL009C	-0.045	-0.038	-0.020
YER059W	-0.044	-0.030	-0.063
YDR429C	-0.044	-0.054	0.034
YMR055C	-0.042	-0.050	-0.060
YDR216W	-0.042	-0.103	-0.003
YBL015W	-0.042	-0.049	-0.004
YLR255C	-0.041	-0.035	-0.003
YDR027C	-0.039	-0.046	-0.007
YKL041W	-0.039	-0.168	0.030
YDR417C	-0.038	-0.045	0.009
YLL007C	-0.038	0.019	-0.033
YKL212W	-0.037	-0.121	0.024
YHR068W	-0.036	-0.048	-0.016
YLR415C	-0.035	-0.042	-0.007
YNL179C	-0.035	-0.013	-0.046
YPR169W	-0.034	-0.121	-0.038
YCR028C	-0.032	-0.068	0.020
YGR114C	-0.025	-0.198	-0.034
YMR058W	-0.006	-0.051	-0.071
YMR125W	-0.002	-0.072	-0.033
YMR047C	0.003	-0.113	-0.079
YMR032W	0.009	-0.101	-0.063
YOR355W	0.012	-0.035	-0.037
YMR060C	0.016	-0.050	-0.078
YMR001C	0.017	-0.109	-0.079
YNL239W	0.018	-0.064	-0.083
YNL257C	0.025	-0.054	-0.033
YKL148C	0.026	-0.107	-0.076
YGR099W	0.029	-0.082	-0.142
YPL208W	0.039	-0.039	-0.111
YGL233W	0.040	-0.157	-0.146

CHAPTER 3: *HYBRID PROTEIN COMPLEXES SHOW PREFERENCES FOR HETEROGENEITY*

Authors: Sam Lancaster, Bryan Fonslow, Navin Rauniyar, Billy Edelman, Judit Villen, John Yates, Maitreya Dunham

Author Contributions:

Bryan Fonslow and Navin Rauniyar performed the mass spectrometry on the hybrid complexes and wrote the mass spectrometry section of the methods, supervised by John Yates.

Billy Edelman performed the SILAC experiment supervised by Judit Villen.

Sam Lancaster created the strains, grew them in labeled media, did the post-mass spectrometry analysis, and the writing, supervised by Maitreya Dunham.

Concept design by Maitreya Dunham and John Yates.

INTRODUCTION

Hybridization occurs readily in budding yeast, which is a common fungus that has a simple mating system, is single-celled, and lacks complex development thus allowing for extremely diverged strains to form viable hybrids. The *Saccharomyces sensu stricto* are the budding yeast that can mate and form a hybrid with *Saccharomyces cerevisiae.* The most diverged strain from *S. cerevisiae* in the *Saccharomyces sensu stricto* is *Saccharomyces uvarum,* which is about 80% similar on both the nucleotide and amino acid levels (Kellis et al., 2003). This nucleotide and amino acid divergence is the same as from humans to mice and chickens, respectively.

Such diverged alleles interacting inside a single cell allows unique insights into cellular properties like *cis*- and *trans*- mRNA regulation (Tirosh et al., 2009) and divergence in chromatin accessibility (Connelly, Wakefield, & Akey, 2014). Two studies have also used this heterogeneity to understand whether or not diverged proteins interact in hybrids. Piatkowska et al., 2013, used mass spectrometry to show that proteins from diverged parental species complex together *in vivo,* and showed that interspecies heterogeneous protein complexes can in some environments confer heterosis. Leducq et al., 2012, used a different method of demonstrating interspecific interactions in hybrid complexes using a *S. cerevisiae x Saccharomyces kudriavzevii* model where *Escherichia coli* dihydrofolate reductase is attached to proteins and reconstituted when proteins from the two different species are in close proximity. Combined, these methods have shown that interspecific protein-protein interactions in hybrids are a generalizable phenomenon.

Protein-protein interactions coevolve so tightly that coevolution can be used to infer protein interactions (Ramani & Marcotte, 2003). Yet extremely diverged proteins, such as those between *S. cerevisiae* and *S. uvarum,* can interact with one another in the hybrid. Since these interactions have been shown to confer heterosis, we wanted to know to what extent heterogeneous complexes form in hybrids. Our goal is to better understand interspecific hybrid complexes in a *S. cerevisiae* x *S. uvarum* hybrid not by qualitatively determining interactions, but by determining which protein is preferred in the hybrid, to what degree it is preferred, if heterogeneous complexes are common, and the residues that are responsible for preferences. To accomplish these goals we isotopically labeled the proteins in yeast with ^{15}N, isolated complexes with coimmunoprecipitation (CoIP), and used the isotopic labeling during mass spectrometry to determine which protein is preferred in the hybrid. We confirmed that interspecific protein

complexing is generalizable, and identified several complexes with preferences for proteins from both *S. cerevisiae* and *S. uvarum*.

METHODS

Strain creation and growth

S. cerevisiae tagged hybrids were made by mating strains from the *S. cerevisiae* Tandem Affinity Purification (TAP) collection to *lys2 S. uvarum* CBS 7001 using selective markers (Ghaemmaghami et al., 2003). Diploid *S. cerevisiae* strains with a single allele TAP-tagged were used to compare to the hybrids. TAP-tagged proteins used were Msh2, Tub4, Utp4, Utp14, Ada1, and Spt7, and were confirmed via PCR using insert-specific primers (Ghaemmaghami et al., 2003). Diploid purebred strains were grown in minimal media containing ^{15}N ammonium sulfate (Cambridge Isotope Laboratories, 98% purity) and the hybrid yeast were grown in minimal media without isotope labeling. Cells were harvested at mid log phase, spun, washed, flash frozen, and then cryo-ground in a Retsch planetary ball grinder PM 100. Cell dust was then mixed in an equal weight, and IPed with IgG labeled magnetic DynaBeads. Beads were washed three times, proteins cleaved using ammonium hydroxide, lyophilized, and then sent for LC MS/MS. Each IP was done in three biological replicates.

S. uvarum TAP labeled proteins were created by PCR amplifying the TAP DNA cassette from the TAP collection, with ~80bp overlap with the 5' portion of each gene and the first 80bp of the 5' UTR, so the cassette would land precisely at the C-terminus of the protein. The same procedure was followed for the *S. uvarum*-tagged proteins as for the *S. cerevisiae*-tagged

proteins, but at 25°C instead of 30°C to accommodate optimal growth conditions. Point mutations were made via overlap extension PCR. After transformation selection of PCR fragment integration used a *URA3* marker that was placed 200bp downstream of the tagged gene on the PCR product.

Mass Spectrometry

Lyophilized protein pellets were resolubilized in 8 M urea and reduced with 5 mM tris(2-carboxyethyl)phosphine (TCEP) for 20 minutes at room temperature and alkylated in the dark with 10 mM iodoacetamide for 20 min prior to digestion. The samples were diluted with 6× volumes of 25 mM Tris-HCl and digested overnight at 37 °C with sequencing grade modified trypsin (Promega). The next day the reaction was stopped by acidification with formic acid at 5% final concentration. The protein digest was pressure-loaded onto a biphasic trapping column packed with 2.5 cm of 5 μm Partisphere strong cation exchanger (HiChrom), followed by an additional 2.5 cm of 5 μm Aqua C18 resin (Phenomenex). The resin-bound peptides were desalted with buffer A (5% acetonitrile/0.1 % formic acid) by flowing through the trap column. The trap and analytical 100 μm i.d. capillary with a 5 μm pulled tip packed with 15 cm of 3 μm Aqua C18 resin (Phenomenex) columns were assembled using a zero-dead volume union (Upchurch Scientific). LC S/MS analysis of the samples was performed on LTQ Orbitrap Velos (Thermo Scientific) interfaced at the front end with a quaternary HP 1100 series HPLC pump (Agilent Technology) using MudPIT technology (EC, 3rd, & L, 2003). Tandem MS spectra were collected in a data-dependent fashion with a repeated cycle of one full-scan MS1 spectrum (300 – 1750 *m/z*) at a resolution of 60,000, acquired using the orbitrap analyzer, followed by 20 data-

dependent MS/MS CID spectra in the linear ion trap analyzer at a 35% normalized collision energy. The maximum ion accumulation times were set to 500 ms for MS1 scans and 25 ms for MS/MS scans. Other instrumental parameters include dynamic exclusion with a repeat count – 1, repeat duration – 30, exclusion list size – 500, and exclusion duration – 120 ms.

The resulting spectra were extracted from the Xcalibur data system format into MS2 format using RawXtract. Protein identification was done with the ProLuCID algorithm by searching MS2 spectra against the database generated by combining the protein sequences of *S. cerevisiae* and *S. uvarum* and filtering it to <1% false positive at the spectrum level using the DTASelect2.0 program. Carbamidomethylation on cysteine was defined as fixed modification and oxidation on methionine was included as variable modification in the database search criteria. Census (Park et al., 2014) was used to obtain quantitative ratios between the ^{14}N and ^{15}N version of each peptide from the result file. Because the *S. cerevisiae* and *S. uvarum* cells were ^{14}N- or ^{15}N-labeled, the peptide ratios between the light and heavy isotopes were directly used as surrogates for protein expression ratios between the two *Saccharomyces* species.

In silico trypsin digest was performed using http://web.expasy.org/peptide_mass/.

Preference Scores

We developed a preference score for each protein that indicates which ortholog is preferred and the strength of the preference. The hybrid was grown in light media (^{14}N) and the purebred was grown in heavy (^{15}N). For each peptide, a light:heavy ratio was determined. We divided each protein into peptides that are unique and not unique between the two species, and then find the average light:heavy ratios for the peptides in each category. Then we take the not

unique:unique ratio to infer the relative amount of each protein present in the hybrid (Figure 1). We subtract 1 from this score and then take the $\log_2$ to generate a preference score. A lower preference score means a *S. cerevisiae* preference and higher means a *S. uvarum* preference (Figure 2). A 95% confidence interval (CI) is drawn from three biological replicates. To prevent errors in multiple hypothesis testing, we created a CRAPome (Mellacheruvu et al., 2013) by first finding the proteins that are part of the union of three replicates, and then identifying proteins that are in more than one union of a complex. Any protein in more than one union is then thrown out of the data, and only the most abundant proteins also a part of the complex are given preference scores.

Preference score: Log_2([not-unique:unique of average light:heavy]-1)

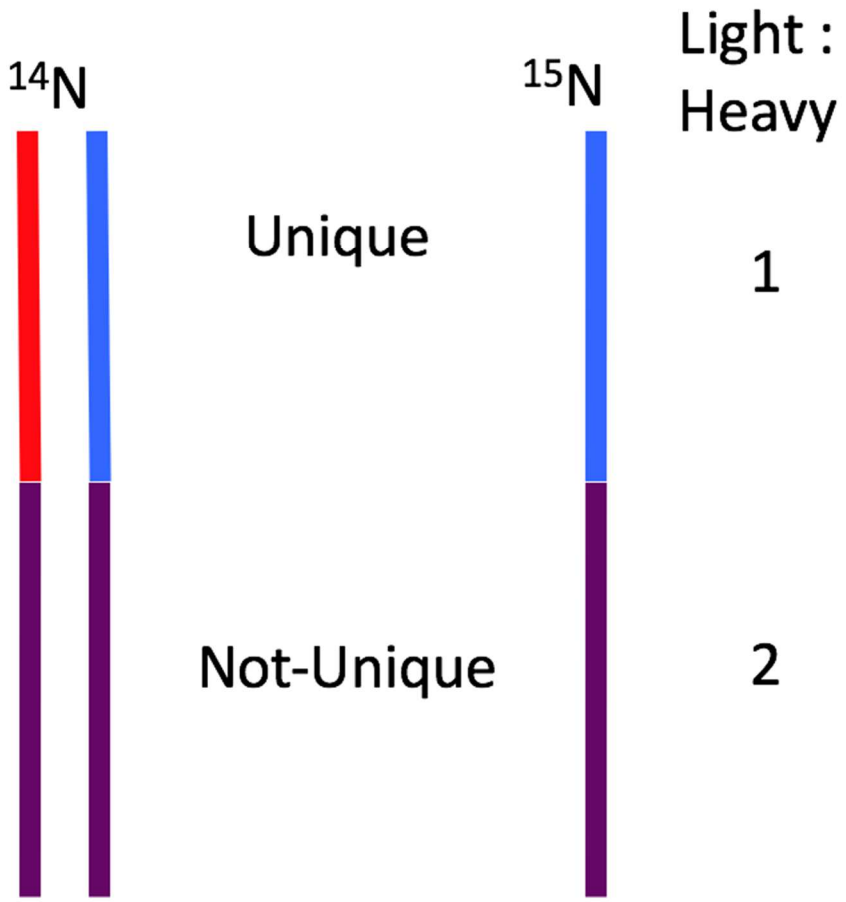

Figure 3.1. Diagram illustrating how preference scores are calculated. Each vertical line split into two colors represents the primary structure of a protein. Each color represents the average of all either the unique peptides (red or blue) or not unique peptides (purple). Both orthologs of any CoIPed hybrid protein are on the left, and the purebred protein is on the right. First the light to heavy ratio is taken between the peptides corresponding to the isotopically labeled purebred, in this case blue. The second light:heavy ratio is between the not-unique peptides. In this example the third ratio of not unique peptides:unique peptides is 2, indicating there is an equal amount of each parental ortholog present in the hybrid. This would be a preference score of 0.

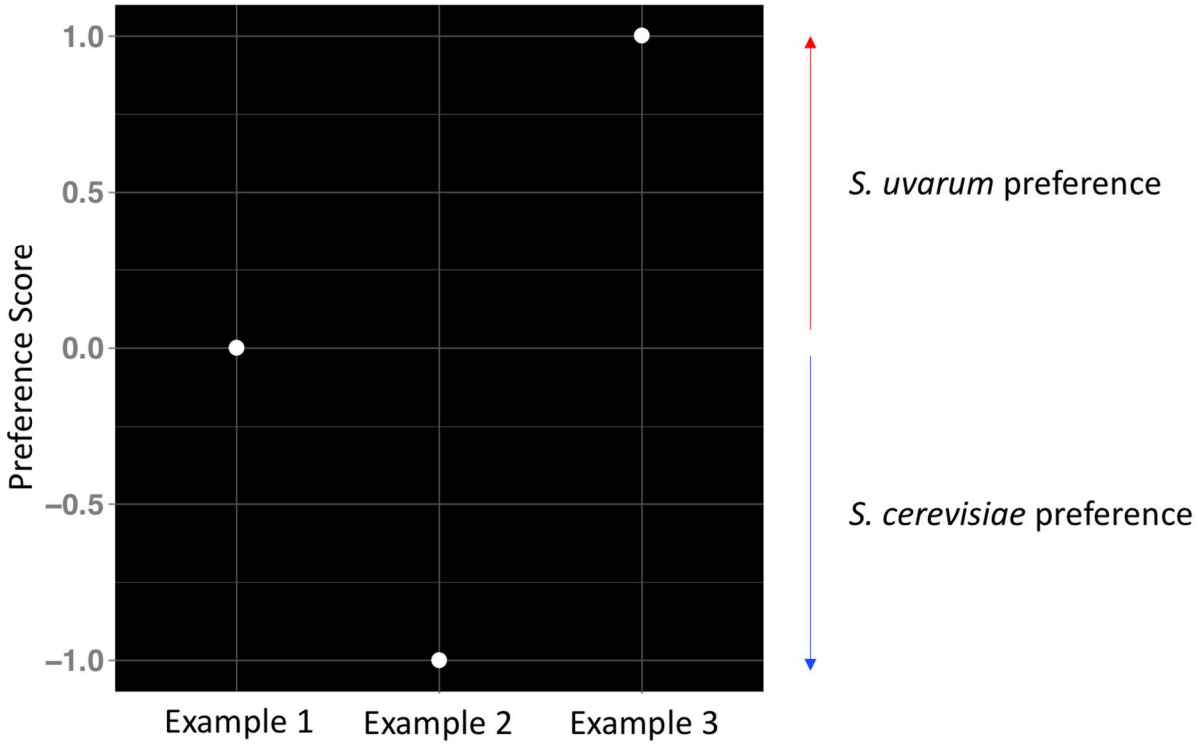

Figure 3.2. Example of the preference scores for any CoIPed hybrid proteins. Example 1 is no preference, Example 2 is twice as much *S. cerevisiae*, and Example 3 is twice as much *S. uvarum.* Example 1 corresponds to Figure 1.

Protein Allele Specific Expression

Protein allele specific expression (pASE) (Khan et al., 2012) was determined for the entire yeast proteome using Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) (Ong et al., 2002; Rappsilber, Mann, & Ishihama, 2007). Isotopically heavy lysine was from Cambridge Isotope Laboratories (U-13C_6, 99%; U-15N_2, 99%; and U-15N2, 99%). The same genetic backgrounds (*S. cerevisiae, S. uvarum,* and hybrid) were used for SILAC as with the tagged complexes. These strains were grown in synthetic complete medium to mid log phase, frozen, the proteome was isolated, mixed 1:1:1, and then mass spectrometry was performed. LC

was run on an nLC easy 1000 from Thermo Fisher with five fractions: flow through, 5%, 10%, 20%, and 40% acetoniritle. Each fraction was analyzed on a Q-Exactive benchtop mass spectrometer (Thermo Fisher) using 120minute gradient via an Easy nLC 1000 HPLC. Samples were loaded onto a 75-µm ID x 30cm micro-capillary column loaded with Reprosil C18 3-µm particles. The gradient conditions were 3-45% acetonitrile in 0.15% formic acid over the course of 120 min. MS spectra were collected in the Orbitrap while MS/MS spectra were collected in the ion-trap by selecting the 20 most abundant ions in a data-dependent manner.

RESULTS

We developed a method to determine the composition of hybrid protein complexes by comparing the hybrids to ^{15}N isotope labeled purebred complexes. Four complexes in particular received our attention: Msh2-Msh3/Msh6 mismatch repair complex, the small ribosomal subunit (SSU) processome, the Spt-Ada-Gcn5 Acetyltransferase (SAGA) Complex, and the gamma-tubulin complex. The Msh2-Msh3/Msh6 complex is a small, abundant complex formed by two sets of dimers with Msh2 that have overlapping but distinct roles (Sia, Kokoska, Dominska, Greenwell, & Petes, 1997). It has also been shown to contribute to genetic incompatibilities in the hybrid (Hunter, Chambers, Louis, & Borts, 1996), making it of particular interest in our model. The SSU processome is a large (80S) ribonucleoprotein complex containing at least 70 proteins and U3 snoRNA that processes the 18S portion of the eukaryotic ribosome (Woolford & Baserga, 2013). Many components of the complex are essential, and any preferences we observe may have implications for the structure of the hybrid ribosome. The SAGA complex acetylates histones and recruits transcription machinery to genes in response to environmental stimuli (Han,

Luo, Ranish, & Hahn, 2014; Sterner & Berger, 2000). One of the fundamental questions about heterosis is how a hybrid uses its genetic complexity to better respond to the environment, and complexes affecting transcription lie at the heart of this problem. The last complex is the gamma-tubulin complex, which is attached to the spindle pole body and nucleates microtubules. We surmised that this complex would be highly sensitive to interspecies differences because three of its members (Spc98, Spc97 and Tub4) are haploinsufficient. The last member is Spc110 – a homodimer, which may decrease the likelihood of forcing interspecific interactions. Together these complexes should give a broad picture of the forces influencing the composition of protein complexes in hybrids.

Proteins in Multiple IPs

Using IPs followed by LC MS/MS on protein complexes in hybrid yeast, we were able to generate our own CRAPome (Mellacheruvu et al., 2013), which represent all of the proteins that bind, whether or not they are tagged, to the antibodies or magnetic beads during IP. Our methods generate a CRAPome that largely overlaps with the published work (730 total *S. cerevisiae* proteins, 108 are specific to this data, and 50 of those were only observed in the hybrids).

Msh2 Complex

We first quantitatively determined the components of heterozygous protein complexes in hybrid yeast with *S. cerevisiae* tagged Msh2. We did three biological replicates with the *S. cerevisiae* ortholog. In this set of experiments, the three main components of the mismatch repair

complex, Msh2, Msh3, and Msh6 from both species, were present in the highest concentrations. In the complex we found that the *S. cerevisiae* copies of Msh2 and Msh3 were preferred. We also find that the *S. uvarum* copy of Msh6 is preferred, even though the *S. cerevisiae* protein is tagged. (Figures 3 and 4)

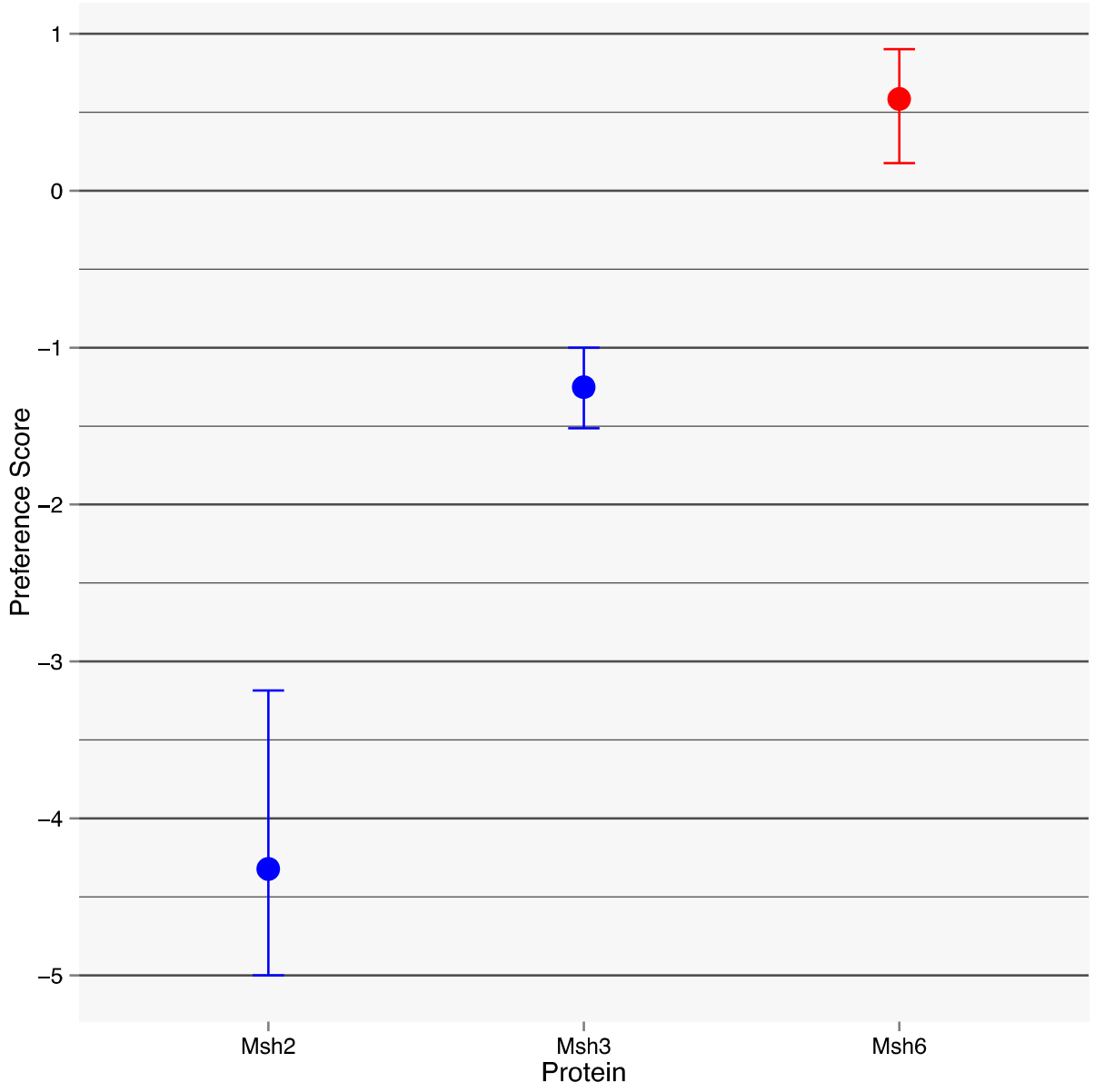

Figure 3.3. Preference scores with 95% CI for the Msh2 complex and its major components. Blue is *S. cerevisiae* preference, purple is no preference, and red is *S. uvarum* preference.

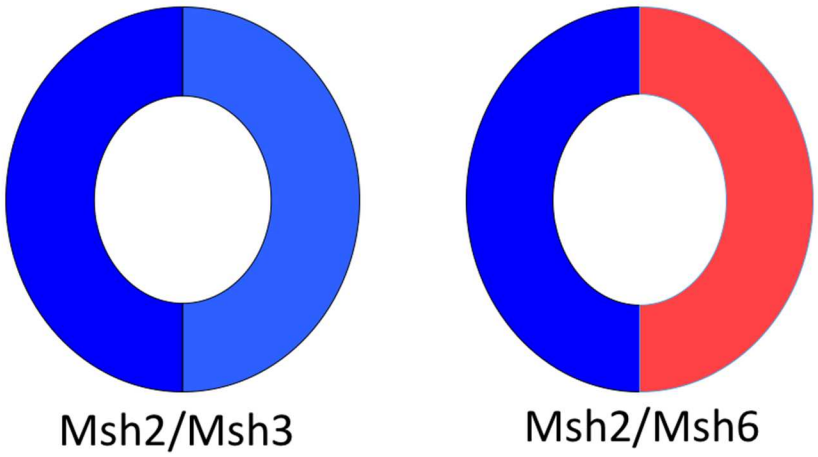

Figure 3.4. Hybrid Msh2 complexes when *S. cerevisiae* Msh2 is tagged. Blue is a *S. cerevisiae* preference and red is a *S. uvarum* preference.

We further identified new interacting partners with the Msh2 complex in the hybrid, for example Pif1 – a DNA helicase involved in DNA damage repair (Supplemental Figure 1). The fifth most common *S. cerevisiae* protein in the hybrid was Rad16, which is also involved in DNA damage repair, and it has been shown to act independently of Msh2 (Fleck, Lehmann, Schär, & Kohli, 1999). Rad16 has a DNA binding partner, Rad7, which was observed at lower levels in our dataset and may not have been suspected to interact with the Msh2 complex if Rad16 had not been abundant. Both Rad16 and Rad7 were present in higher concentrations in the hybrid than the purebred. The fifth most common protein overall was a *S. uvarum* protein that lacked a *S. cerevisiae* homolog and was annotated as Sbay_5.223, on www.saccharomycessensustricto.org (Scannell et al., 2011). These may indicate new interactions, and the first annotated function for the unidentified *S. uvarum* protein.

SSU Processome

The ribosome plays an integral part in cellular physiology, and we applied our methods on the *S. cerevisiae* Utp4 and Utp14 components of the ribosomal SSU processome. We further TAP-tagged *S. uvarum* Utp4 to see the effect of orthologous IPs. With *S. cerevisiae* Utp4 tagged, no *S. uvarum* Utp4 CoIPed. When *S. uvarum* Utp4 was IPed, only one *S. cerevisiae* Utp4 peptide was identified. In the *S. cerevisiae* Utp4 IP, there are two other proteins for which we only see no *S. uvarum* peptides: Utp14 and Utp20. (Supplemental Figures 3, 4, and 5)

Utp4 is a member of the UtpA subcomplex of the SSU processome, yet we were able to identify members of other subcomplexes of the processome, including some unclassified members, like Utp14. The location of Utp14 is unknown, and the members of the processome that CoIP in most abundance with it are Utp6 and Utp12, both of which are members of the UtpB subcomplex (Phipps, Charette, & Baserga, 2011). This suggests that Utp14 interacts more closely with this subunit, and in a heterogeneous way because these members show a preference for the *S. uvarum* allele.

Gamma-tubulin Complex

We purified the gamma-tubulin complex from hybrid yeast by TAP-tagging the *S. cerevisiae* copy of Tub4, *S. cerevisiae* Spc110, and *S. uvarum* Tub4. These IPs yielded only four proteins whose preferences were abundant enough to quantify: Tub4, Spc98, Spc97 and Spc110, although many members of the spindle pole body were identified.

Tub4, which was tagged in both parental species, showed no statistically significant preference in any instance. These findings are consistent with the ring-like structure of the

gamma-tubulin complex, where there are at least 13 copies of Tub4 present. Furthermore, since we identify the spindle pole body, and there are many gamma tubulin complexes attached to it, there might be multiple copies of the gamma tubulin complex for every tagged copy. Together these effects would dilute out any preference for the tagged copy of Tub4.

The only preference in the gamma tubulin complex is for *S. uvarum* Spc98. When *S. uvarum* Tub4 is tagged, we only see the *S. uvarum* copy of Spc98 present (Supplemental Figure 6). Even when the *S. cerevisiae* Tub4 is tagged, we see a preference for *S. uvarum* Spc98 (Figure 5). These findings indicate a strong preference for Spc98 in the hybrid gamma tubulin complex, perhaps due to stronger binding to Tub4 (Figure 6). Contributing to this phenomenon is a residue in the gamma-tubulin binding domain of Spc98 that is phosphorylated in *S. cerevisiae* (S675) and phosphomemetic (D) in *S. uvarum* (T.-C. Lin et al., 2011). This is the only *S. cerevisiae* phosphorylated residue that is different between the two species. When the *S. uvarum* residue is mutated to the *S. cerevisiae* copy, the binding strength of the *S. uvarum* protein increases >4 fold.

When using the *S. cerevisiae* copy of Spc110, the binding strength of Spc110 was not strong enough to quantify members of the spindle pole body or the gamma tubulin complex. There was a clear preference for the *S. cerevisiae* copy of Spc110, although there is *S. uvarum* Spc110 present. The presence of the *S. uvarum* copy indicates that there might be both parental copies binding to each other in the homodimer. It is possible that there are multiple gamma-tubulin complexes in every IP, as previously mentioned, which could be an alternate explanation to why the *S. uvarum* copy is identified. If orthologous proteins are forming a homodimer, that would highlight the robustness of interspecific protein interactions.

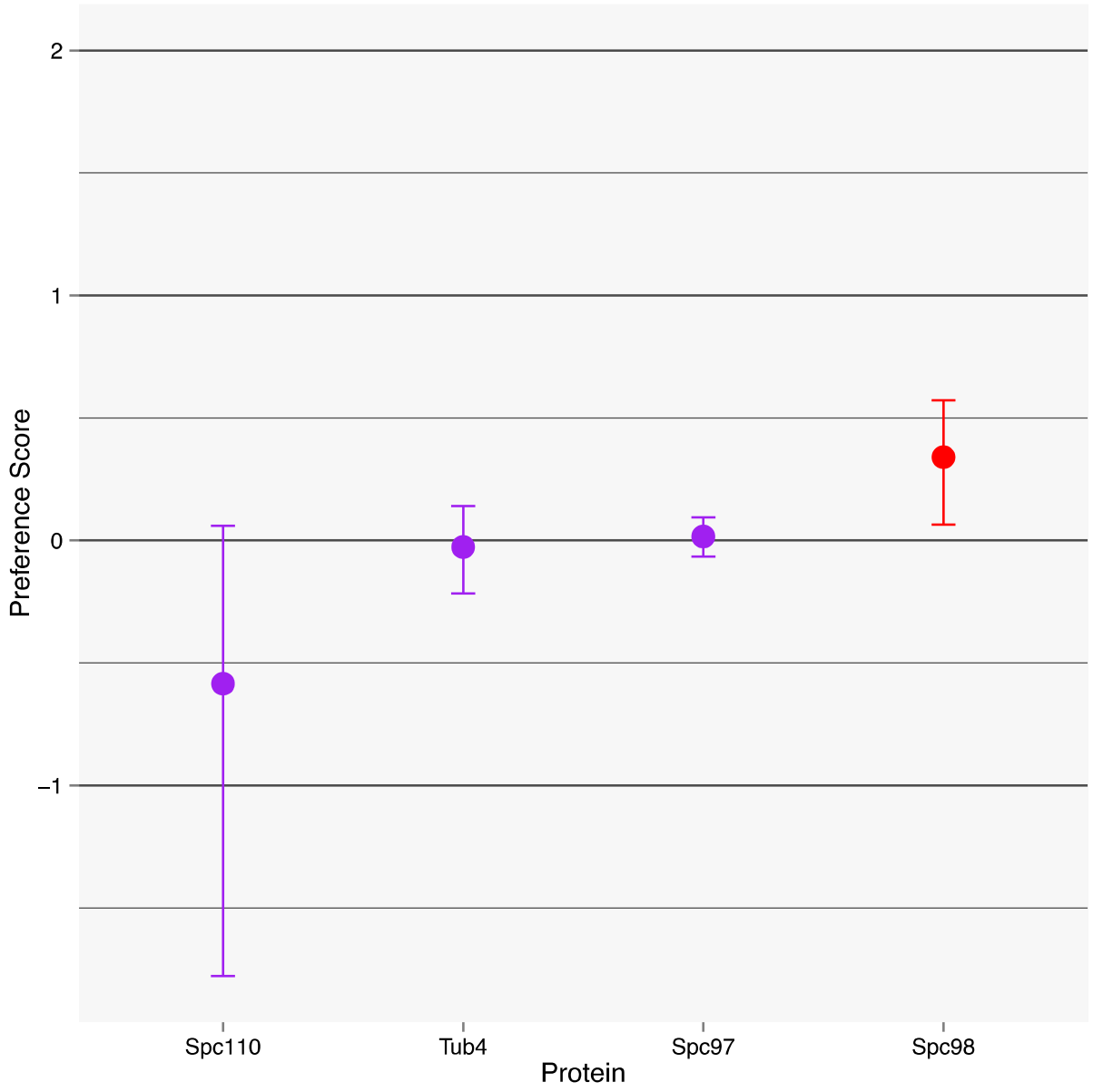

Figure 3.5. Top CoIP proteins that are also part of the gamma tubulin complex when *S. cerevisiae* Tub4 is TAP-tagged. Blue is *S. cerevisiae* preference, purple is no preference, and red is *S. uvarum* preference. Error bars are 95% CI.

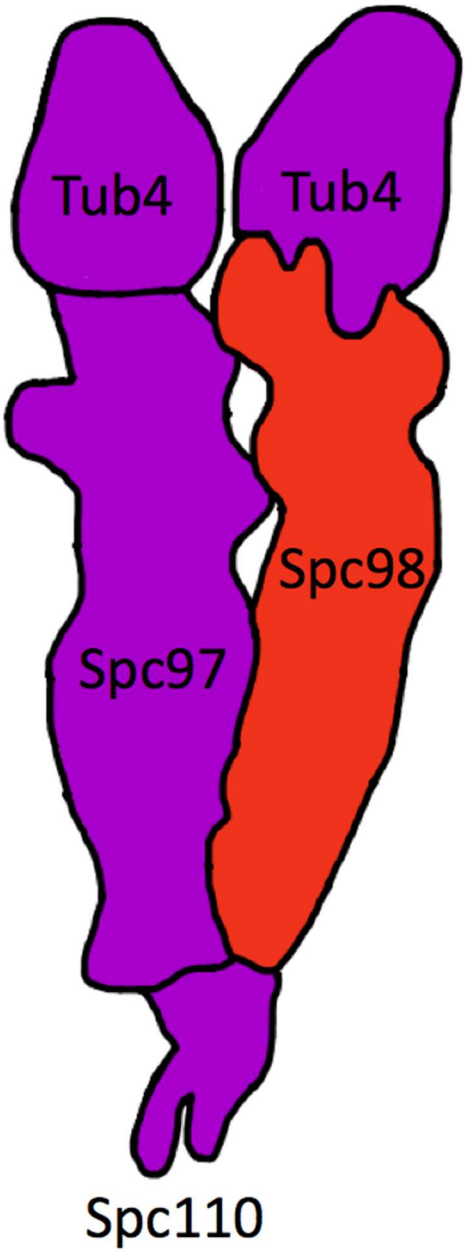

Figure 3.6. Hybrid gamma tubulin complex. Purple is no preference and red is a *S. uvarum* preference.

SAGA Complex

We performed two IPs from the SAGA complex: *S. cerevisiae* Ada1 and Spt7, which identified mostly overlapping proteins. There were, however, slight differences in preferences between the two. Spt8 had a very small preference for the *S. uvarum* copy in the Ada1 IP, though showed no preference in the Spt7 IP. If there is any statistical preference for Spt8, it must be very small. Sgf29 showed no preference in the Ada1-tagged IP and it showed a preference for *S. cerevisiae* in the Spt7 IP. According to SGD, *S. cerevisiae* Sgf29 has more interactions with Spt7 than with Ada1, suggesting stronger intraspecific interactions in the hybrid. Other proteins identified include other members of the SAGA complex, such as Taf9, which has more identified physical interactions with Spt7 than Ada1, and Ngg1 (alias: Ada3). In the Ada1 IP, we were able to quantify Ubp8 that was not in the Spt7 sample. From these IPs together, we can build a model of the hybrid SAGA complex (Figure 7). (Supplemental Figures 2 and 3)

In the Ada1 IP, we saw no preference between the *S. cerevisiae* and *S. uvarum* copy of Spt7; however, when the *S. cerevisiae* copy of Spt7 was TAP-tagged, it was the only ortholog present. These findings strongly suggest that only a single copy of Spt7 is present in the SAGA complex.

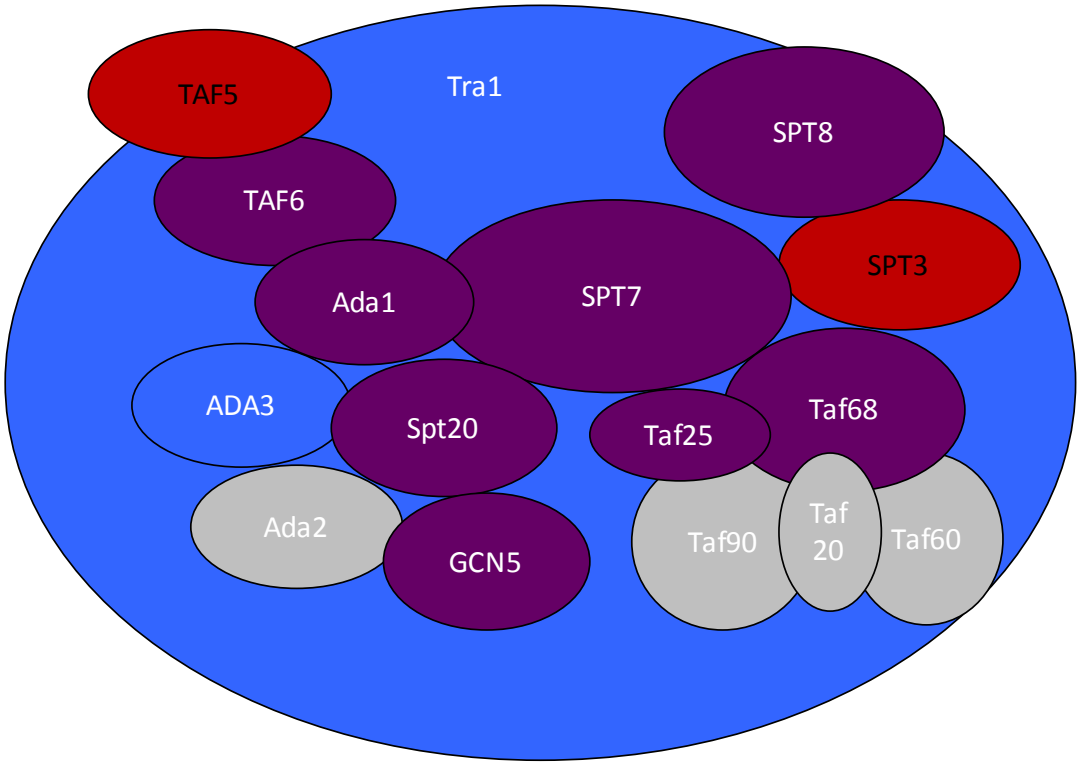

Figure 3.7. Hybrid SAGA Complex. Purple represents no preference between either ortholog, blue is a *S. cerevisiae* preference, red is a *S. uvarum* preference, and gray are members of the complex for which there is not enough data to create a preference score. Adapted from Sterner and Berger, 2000.

Protein Expression

Inevitably protein expression will influence preferences in protein complexes. To measure protein expression, we used SILAC on both purebreds and the hybrid. From these data we were able to calculate pASE (Khan et al., 2012). Higher pASE indicates more relative *S.*

cerevisiae to *S. uvarum.* There is a negative association between ortholog preference and ortholog expression, although the correlation is not strong. (Figure 8)

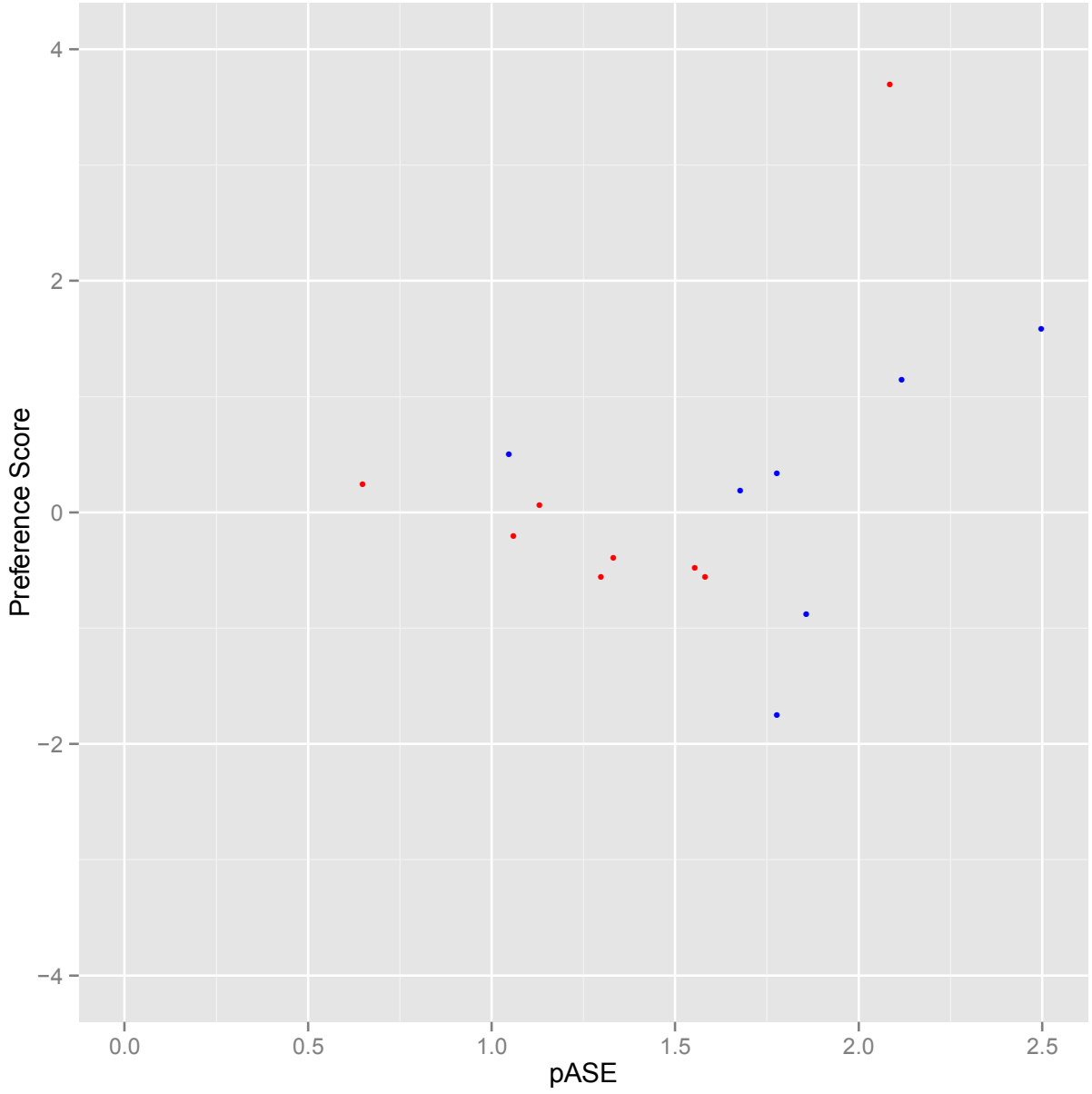

Figure 3.8. Protein score plotted against pASE as determined via SILAC from *S. cerevisiae* Utp4 (red), and Msh2 (blue). $R^2 = 0.15$

Variance in the data

The ability to determine preference scores with narrow confidence intervals, which reflects the variability between replicates, is highly dependent on the abundance of the complex. For example, in the gamma tubulin complex, for which there are <600 copies per cell, some components have large confidence intervals for the preference scores. Furthermore, we were not able to quantify any preferences for the spindle pole body, which is attached to the gamma tubulin complex, and for which there are at most two copies per cell. Conversely, the SAGA complex and the SSU processome are abundant, and many members of the complex are identified, although some weakly associated members are not abundant enough to quantify. These results suggest that more abundant complexes yield more meaningful data for these scores. Furthermore, our methods for IPing complexes also have not been optimized for *S. uvarum,* so this will increase the noise when that ortholog is TAP-tagged.

There is variance in the light to heavy ratio for individual peptides as well. For example, in the three biological replicates of Msh2, the statistical variance between the not unique peptides is 0.046, 0.026 and 0.131 (averages 1.63, 1.66, and 2.30); and the variance of the *S. cerevisiae* specific peptides is 0.018, 0.046, and 0.103 (averages 1.57, 1.51, and 2.27). For Msh6, which CoIPed with Msh2, the variance between the not unique peptides is 0.317, 0.118, and 0.406 (averages 2.36, 1.94, and 3.67); and the variance of the *S. cerevisiae* specific peptides is 0.012, 0.323, and 0.052 (averages 0.848, 0.881, and 1.27). Variance will increase for lower abundance proteins for which there are fewer peptides identified; however, for the most abundant members of the complex, like the examples for the Msh2 complex, these are representative of the variance.

Proteins with large variances may have false negative preference scores. Dim1 from the Utp14 IP only had both unique and not unique peptides present in two of the three replicates. In these two replicates there were only one of each type of peptide present. The unique peptides had light to heavy ratios of 0.96 and 0.89, and the not unique peptides had light to heavy ratios of 1.01 and 2.5, respectively. Hence the variance of this preference scores spans negative infinity (all *S. cerevisiae*) to just above 0 (no preference). Such proteins may in fact have preferences for parental orthologs, and to determine these they should be tagged or tighter binding partners should be tagged to generate more peptides.

DISCUSSION

We sought to determine the allelic composition of protein complexes in hybrids and any quantitative preferences for orthologous proteins. To accomplish this, we grew purebred yeast in ^{15}N labeled media to compare to the hybrid, each with the same component of a protein complex TAP-tagged. Samples were mixed, purified, and sent for LC MS/MS. This method has advantages over SILAC because it does not require an auxotrophy. However, because this method uses heavy ammonium sulfate, there are multiple isotopes incorporated in every peptide. With 98% purity, this means that some significant portion of peptides will incorporate a normal-weight nitrogen, making the deconvolution of the heavy peptides from the mass spectra more challenging. In SILAC, every peptide contains a single heavy amino acid, alleviating this problem.

This method is also powerful to understand stoichiometry of protein complexes. In the yeast SAGA complex, there is one copy of Spt7. When Ada1, believed to be a direct binding partner of Spt7 (Han et al., 2014), is tagged, we found that there is no preference in the complex

for either parental copy of Spt7. However, when *S. cerevisiae* Spt7 is tagged we only find the *S. cerevisiae* copy – a finding strongly suggestive of one Spt7 being present in the complex. Conversely, we know that there are many Tub4 copies in the gamma-tubulin complex and spindle-pole body. When either allele of Tub4 is tagged we see no ortholog preferences – a finding which is consistent with the known stoichiometry of this complex.

In the Msh2-Msh3/Msh6 complex and SAGA complex, we found that when the *S. cerevisiae* components are tagged, there were preferences for *S. uvarum* proteins. The underlying mechanism conferring these preferences remains unknown, but heterogeneous complexes such as these have been shown to contribute to heterosis (Piatkowska et al., 2013).

Our findings also suggest important motifs for protein binding. When *S. cerevisiae* Tub4 was tagged, there was a preference for *S. uvarum* Spc98, and we found only *S. uvarum* Spc98 when the *S. uvarum* Tub4 was TAP tagged. The cell uses phosphorylation to nucleate microtubules (Johmura et al., 2011; Vogel et al., 2001), likely by increasing the binding affinity between components of the gamma tubulin complex. There is one *S. cerevisiae* Spc98 phosphorylation residue that is not in *S. uvarum* (T.-C. Lin et al., 2011), and this residue is phosphomemetic in *S. uvarum.* When we changed this to the *S. cerevisiae* form, *S. uvarum* Spc98 binds four times stronger, suggesting the *S. cerevisiae* residue serves to strengthen the binding of the gamma tubulin complex.

The composition of hybrid complexes may elucidate new protein interactions. For example, we identified an abundant, uncharacterized *S. uvarum* protein that binds to the Msh2-Msh3/Msh6 complexes in hybrids. Because of its strength of binding and because it was only found bound to this complex, this protein may play a role in DNA damage repair. Furthermore, we demonstrate hybrid specific interactions between Msh2 and other DNA repair proteins, such

as Rad16, which had been thought to act independently of each other. These interactions may indicate hybrid-specific DNA damage repair mechanisms where both proteins are recruited.

Here we have shown that the processome in hybrids contains members from both species, possibly indicating heterogeneous ribosomes. Translational divergence has been observed in *Saccharomyces,* much of it acting *trans* (Artieri & Fraser, 2014). Such translational divergence likely is related to the structure of the ribosome, and perhaps the heterogeneous nature of the ribosome influences the additive effects of protein expression in the hybrids (Blein-Nicolas et al., 2015). Sequencing the rRNA to determine its composition in the hybrid would be an important step to understanding the ribosome and how proteins are translated in the hybrid. Perhaps sequence read abundance of rRNA will correlate with preference scores of proteins that are known to act on those segments.

In addition to understanding the molecular forces that influence hybrid protein-protein interactions, we looked at ortholog-specific protein expression using pASE, and there was no apparent association between expression and ortholog preference. Certainly, one cannot rule out that expression influences protein complex composition, and selectively altering expression for an ortholog would necessarily alter the composition of a complex. Future studies should compare high quality RNAseq data to preference scores to see how gene expression is associated with protein complex composition. Further, in order to find out at what point changes in expression begin to affect composition, one could attach a member of a complex to different strength promoters to see at what point dosage affects preference scores.

Since protein complexes are responsible for virtually every cellular process, the underlying causes of heterosis must partially lie in the composition of the protein complexes. Here we have developed a method and shown quantitative preferences for four complexes in a *S.*

cerevisiae x *S. uvarum* hybrid. These preferences indicate heterogeneous complexes are preferred, and these have been shown to contribute to heterosis. We have also shown new protein interactions in the hybrid, indicating hybrid-specific protein interactions. We further believe that we have identified a single phosphorylated amino acid residue in Spc98 that strongly influences preference scores in the gamma tubulin complex. This method may be extended to other instances where significantly different alleles interact, including viruses interacting with different host organisms or in cases of disease-causing haploinsufficiency.

SUPPLEMENT

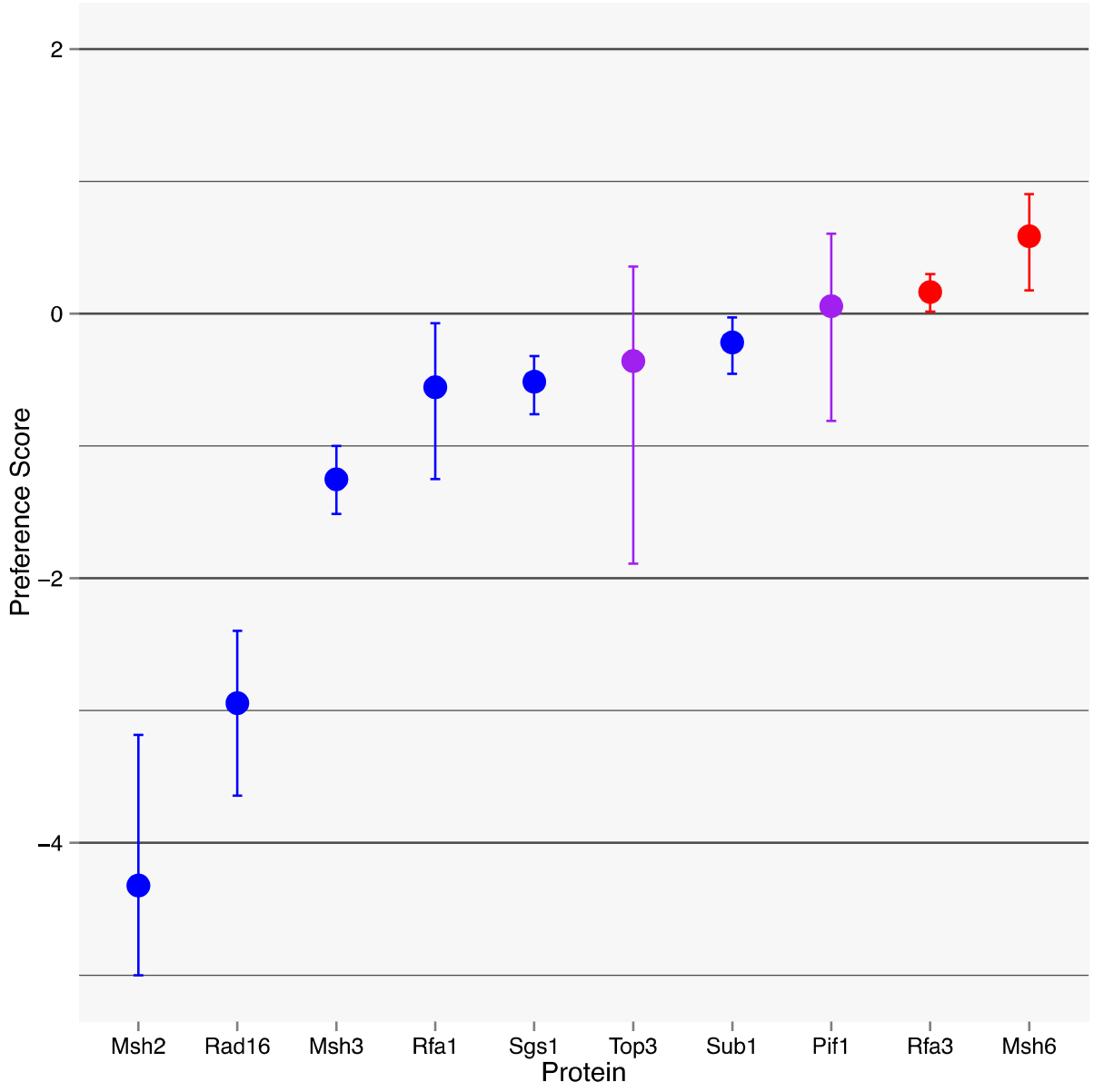

Figure 3.9. Top CoIP proteins that are also a part of DNA damage repair when *S. cerevisiae* Msh2 is TAP-tagged. Blue is *S. cerevisiae* preference, purple is no preference, and red is *S. uvarum* preference. Error bars are 95% CI.

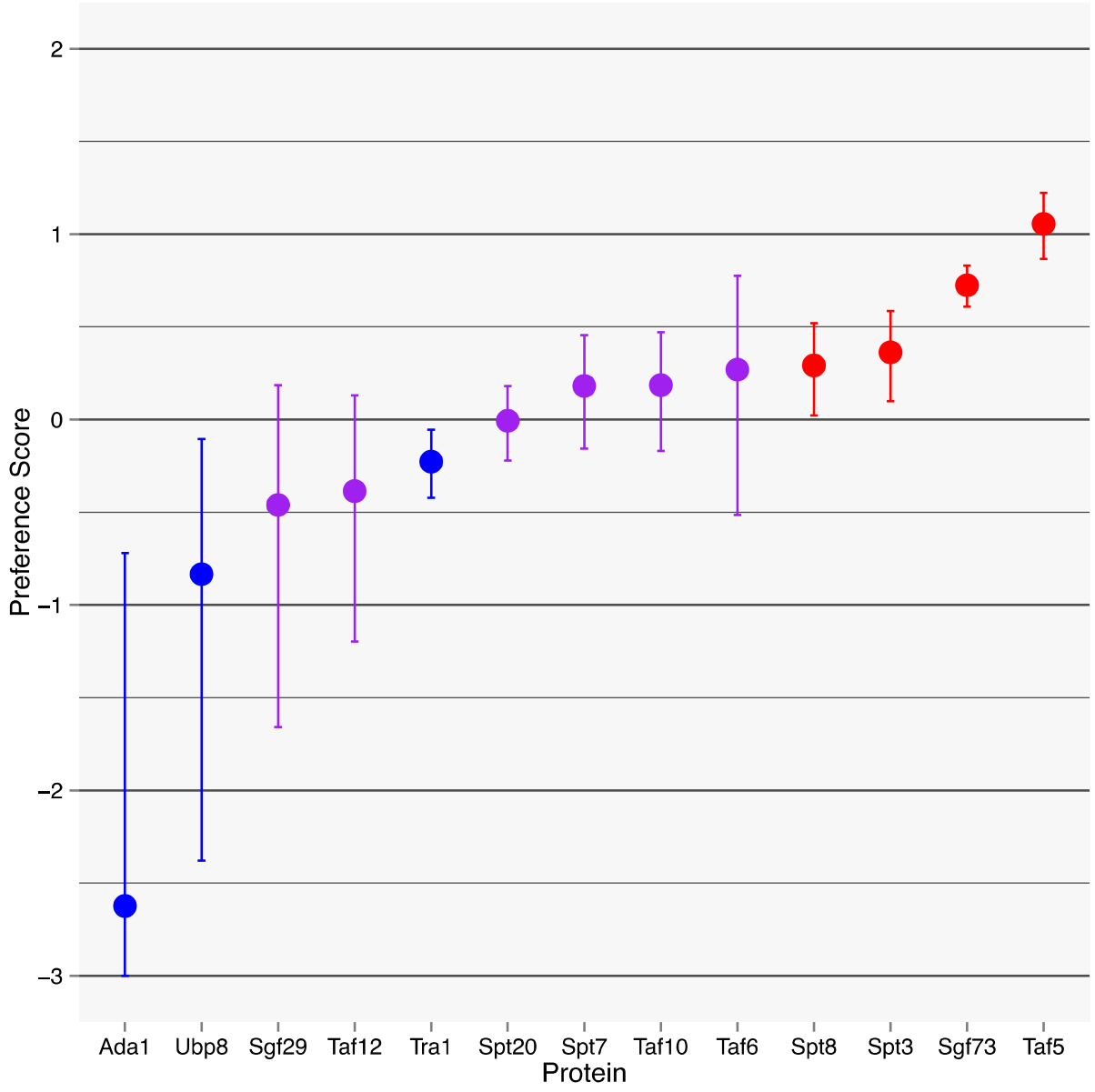

Figure 3.10. Top CoIPed proteins that are also part of the SAGA complex when *S. cerevisiae* Ada1 is TAP-tagged. Blue is *S. cerevisiae* preference, purple is no preference, and red is *S. uvarum* preference. Error bars are 95% CI.

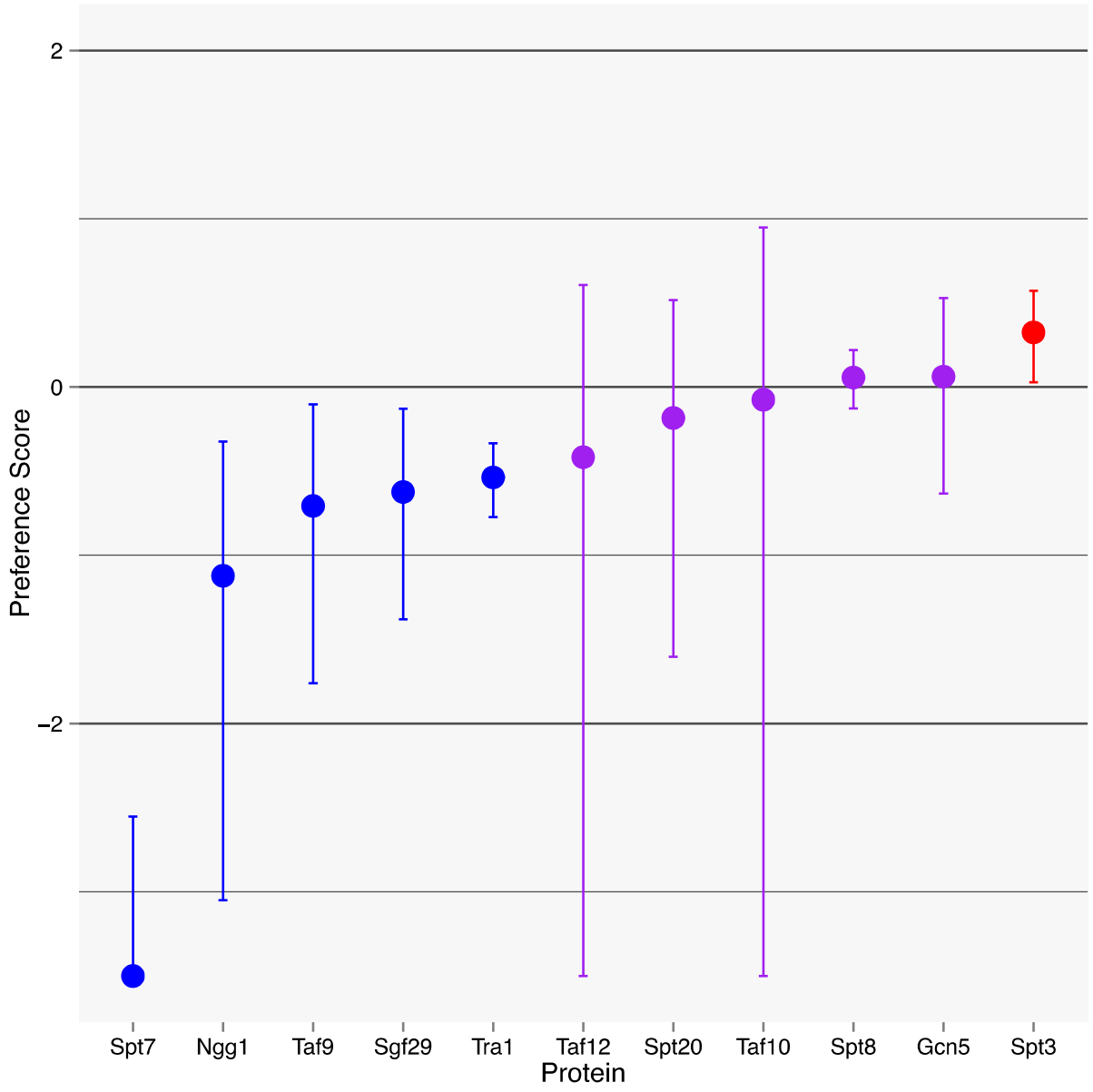

Figure 3.11. Top CoIP proteins that are also part of the SAGA complex when *S. cerevisiae* Spt7 is TAP-tagged. Blue is *S. cerevisiae* preference, purple is no preference, and red is *S. uvarum* preference. Error bars are 95% CI.

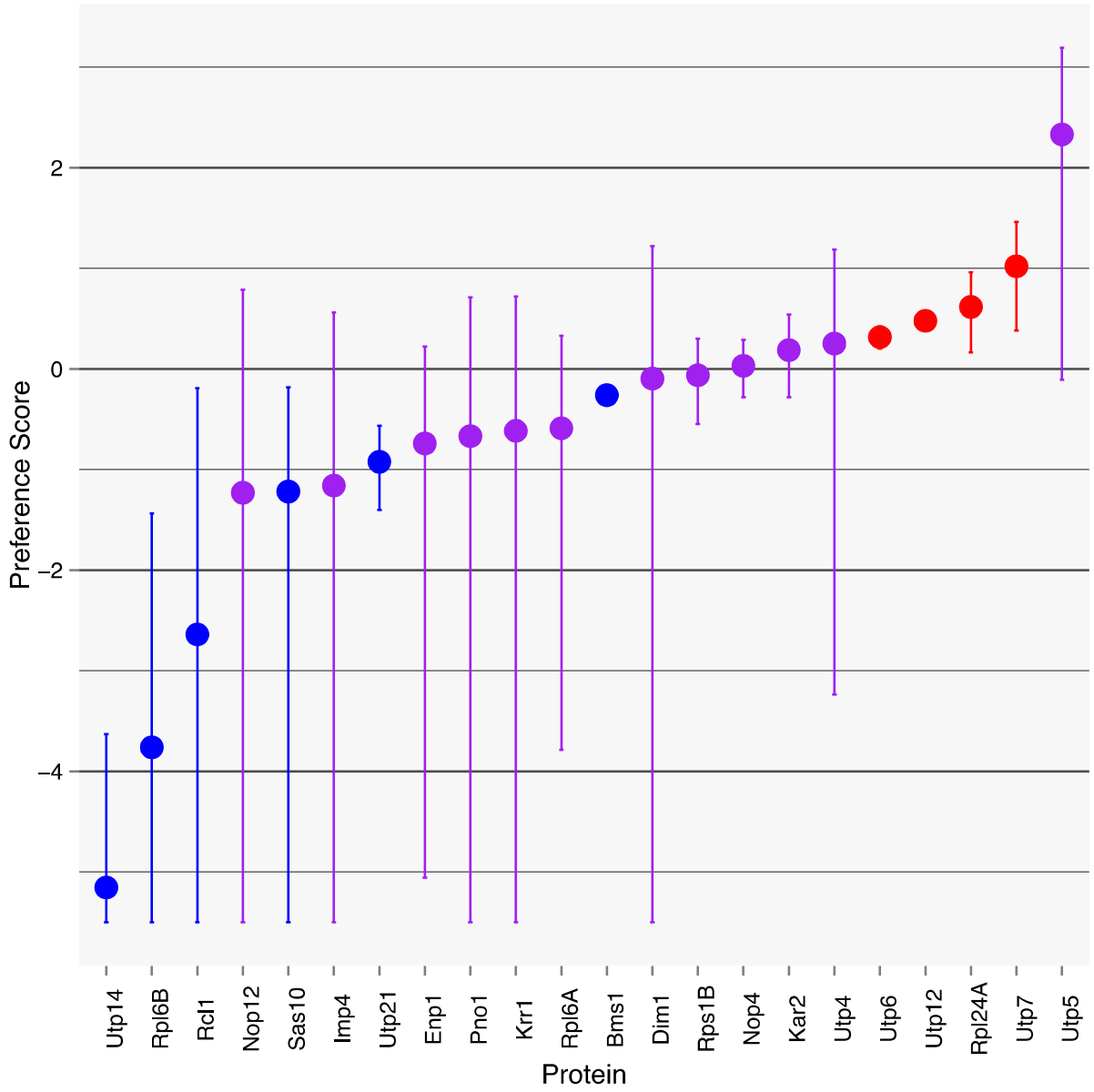

Figure 3.12. Top CoIP proteins that are also part of the SSU processome when *S. cerevisiae* Utp14 is TAP-tagged. Blue is *S. cerevisiae* preference, purple is no preference, and red is *S. uvarum* preference. Error bars are 95% CI.

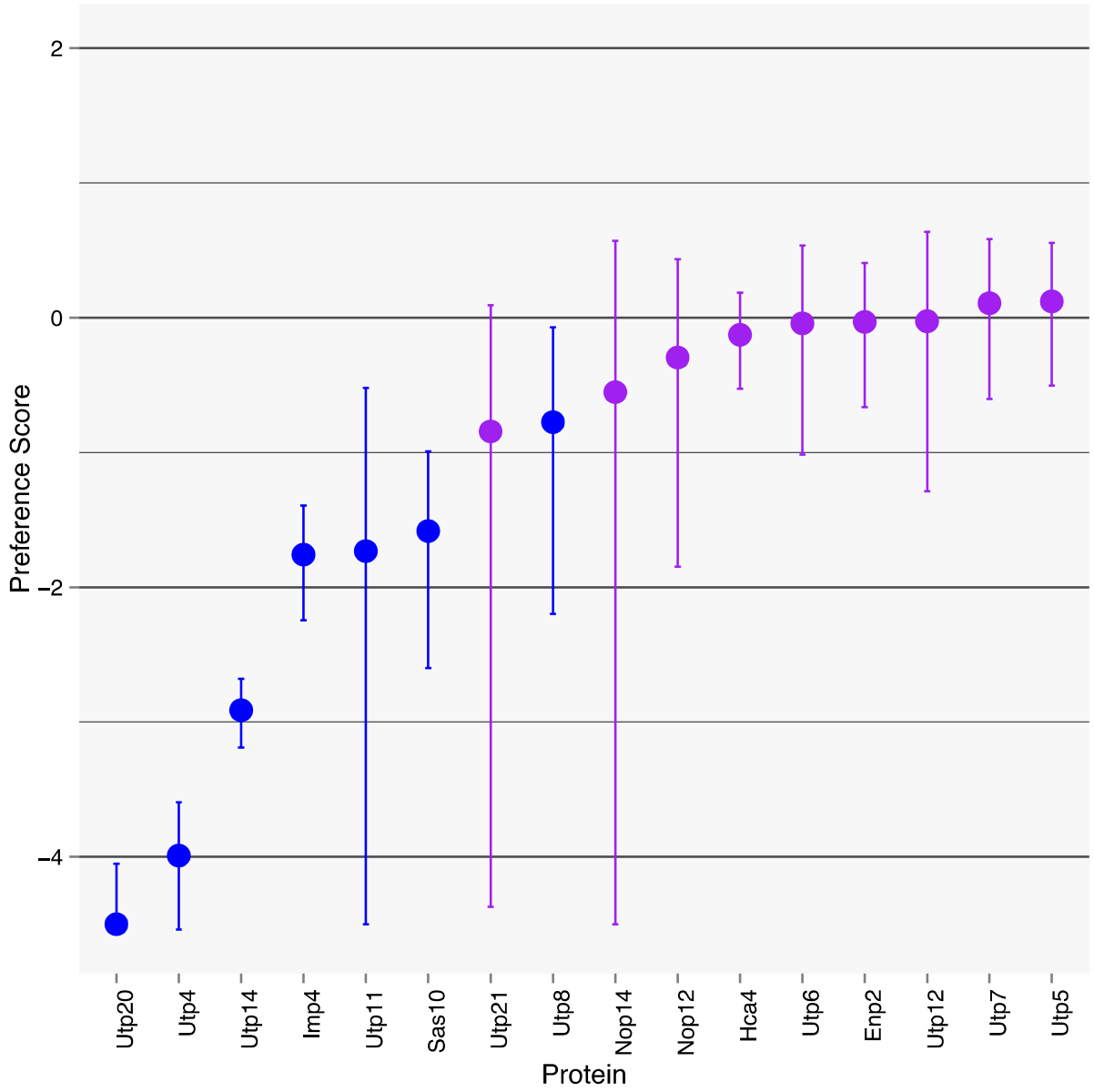

Figure 3.13. Top CoIP proteins that are also part of the SSU processome when *S. cerevisiae* Utp4 is TAP-tagged. Blue is *S. cerevisiae* preference, purple is no preference, and red is *S. uvarum* preference. Error bars are 95% CI.

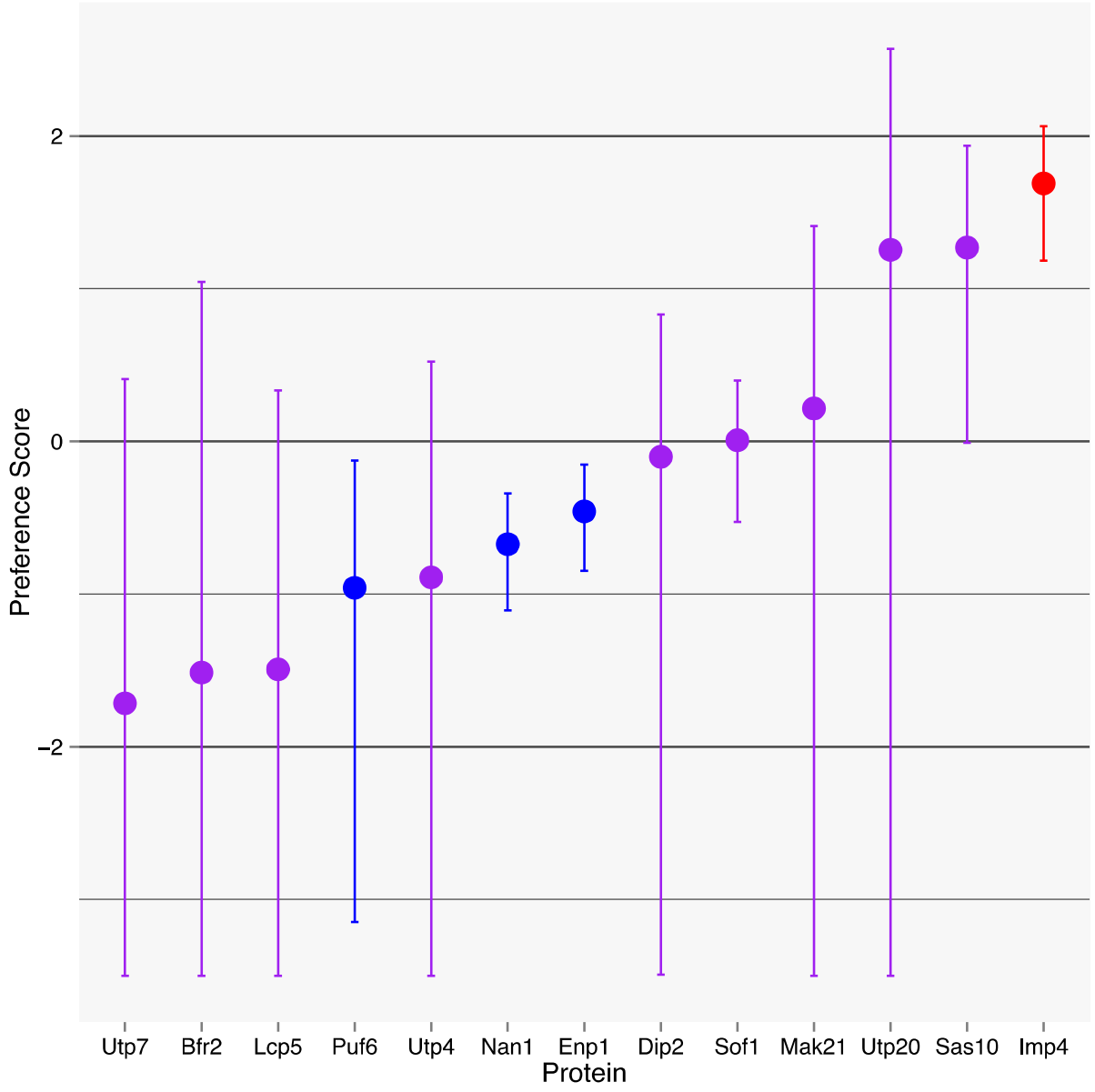

Figure 3.14. Top CoIP proteins that are also part of the SSU processome when *S. uvarum* Utp4 is TAP-tagged. Blue is *S. cerevisiae* preference, purple is no preference, and red is *S. uvarum* preference. Error bars are 95% CI.

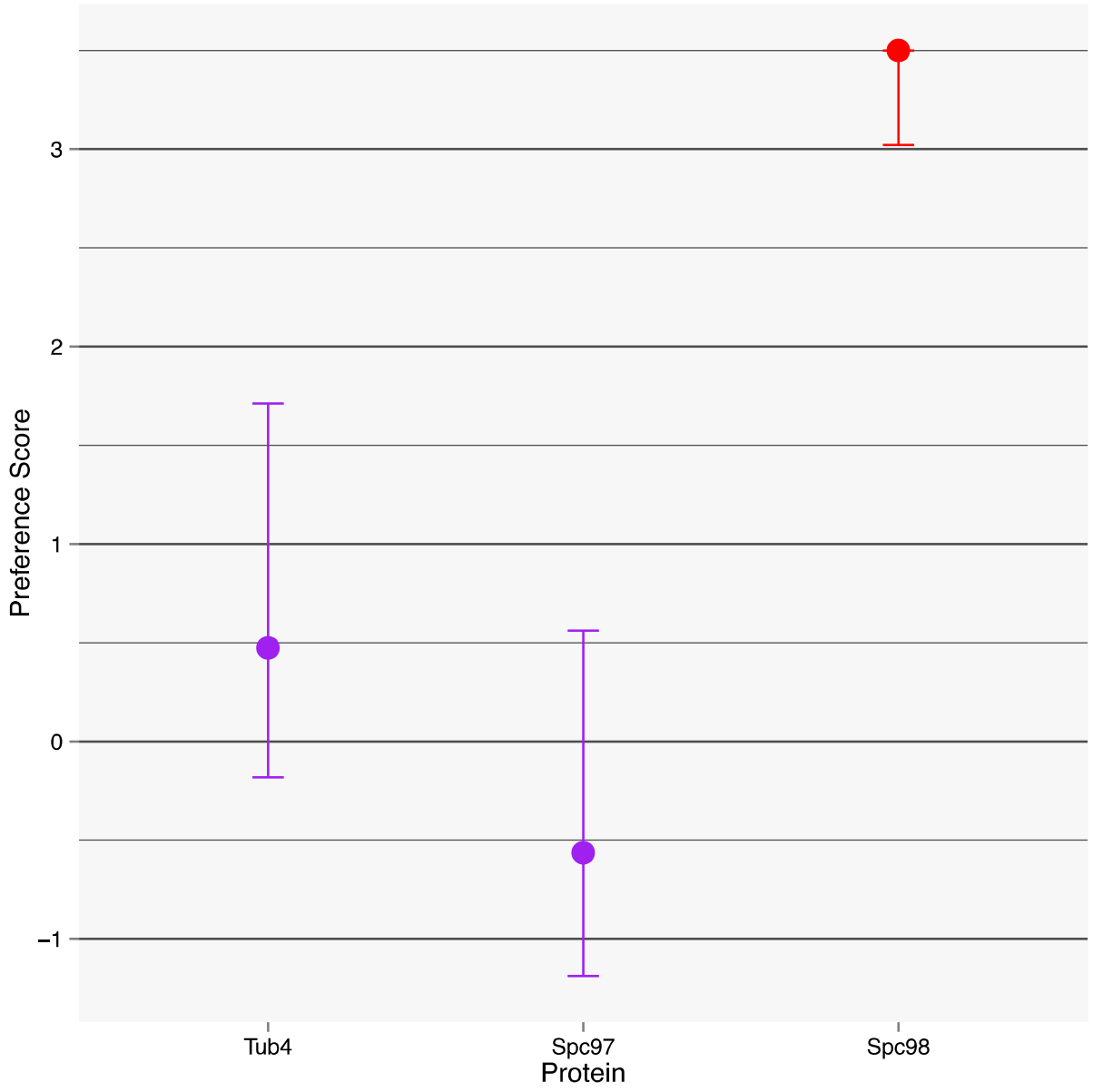

Figure 3.15. Top CoIP proteins that are also part of the gamma tubulin complex when *S. uvarum* Tub4 is TAP-tagged. Purple is no preference, and red is *S. uvarum* preference. Error bars are 95% CI.

CHAPTER 4: MATING IN INTERSPECIFIC HYBRIDS

Authors: Samuel Lancaster, Noah Hanson, and Maitreya Dunham

Author Contributions:

Samuel Lancaster did the *en masse* mating experiment, concept design, and writing, supervised by Maitreya Dunham.

Noah Hanson did the individual mating experiments, supervised by Samuel Lancaster and Maitreya Dunham.

Maitreya Dunham contributed to the concept design.

INTRODUCTION

Understanding speciation via reproductive isolation is of paramount importance to solving problems remaining within evolution. Reproductive isolation has been studied in many organisms, including the budding yeast *Saccharomyces cerevisiae*. This organism straddles two important means of evolution: asexual evolution, which drives phenomena such as antibiotic resistance, and natural selection via sexual recombination. Furthermore, *S. cerevisiae* is an excellent model organism with a fast doubling time and a well-developed genetic system. These factors make it excellent to study barriers to genetic recombination. The sexual cycle of *S. cerevisiae* involves emitting pheromones that eventually help yeast find their mating partner. After detection of a pheromone, a protrusion driven by actin then contacts the other cell, fuses with it, and finally fuses the nuclei to form a zygote in a process that is highly complex for a single-celled organism. (Merlini et al., 2013)

S. cerevisiae can mate with a large number of other *Saccharomyces* species, together known as the *Saccharomyces sensu stricto.* Using this clade several studies have tried to understand the forces governing reproductive isolation, with varying success. Low spore viability is a primary cause of reproductive isolation between *S. cerevisiae* and *Saccharomyces uvarum,* which are the most distantly-related organisms in the *sensu stricto* (Chambers, Hunter, Louis, & Borts, 1996). Several nuclear-mitochondrial incompatibilities have been identified between these two contributing to low spore viability, and indicating that such incompatibilities represent a generalized source of reproductive isolation (Chou et al., 2010; Lee et al., 2008). Nonetheless, viable hybrid lineages can be formed, and parents of comparable divergence spawned *Saccharomyces pastorianus* – one of the most widely used brewing yeast.

Prezygotic reproductive barriers have also been studied in the *Saccharomyces sensu stricto,* although not as extensively. It has been shown that strains of *S. cerevisiae* prefer to mate with itself over *Saccharomyces paradoxus,* although *S. paradoxus* shows no preference for intraspecific mating (Maclean & Greig, 2008). Further, those two species have different germination times, suggesting allochronic contributions to reproductive isolation (Murphy & Zeyl, 2012). If strong enough, prezygotic barriers to hybridization could be a complete way to isolate species. To investigate these, we mated the majority of the *S. cerevisiae* deletion collection to *S. uvarum,* determined their mating efficiency, and found several strains with decreased interspecific mating.

METHODS

We inoculated a 12 hour overnight culture of the pooled haploid *S. cerevisiae* deletion collection, which was prototrophic for lysine and methionine and auxotrophic for uracil and leucine, and grew it well shaken at 30°C. An overnight of *lys2 S. uvarum* was also grown. In the morning, the OD_{600} for these overnights was taken, and they were mixed in a 10:1 ratio of *S. uvarum:S. cerevisiae*. A genome preparation was performed on excess deletion collection overnight culture. The mixed cells were then incubated for 5 hours at room temperature to allow mating. After incubation, the cells were plated on –lys –ura synthetic media to select for diploids. These plates were grown for two days, scraped, mixed well, and a genome prep was performed. To compare to intraspecific mating, the same procedure was followed with *lys2 S. cerevisiae* as the mating partner.

The uptag DNA barcode from the deletion cassette was PCR amplified from the samples before and after mating. These barcodes were sequenced on an Illumina Nextseq. We normalized the barcodes to total sequencing reads, took the ratio of sequencing reads from before and after for each strain, and finally the $\log_{10}$ of this ratio was taken to generate a pseudo mating score. The *S. uvarum* mating partner was *lys2* CBS 7001, and the *S. cerevisiae* control was *lys2* FY.

RESULTS

In order to understand potential barriers to gene flow, we investigated the genetic forces that influence mating between *S. uvarum* and *S. cerevisiae.* First, we established the baseline mating efficiency for these species using the protocol from Leu and Murray, 2006 (Leu &

Murray, 2006). The interspecific mating efficiency is 0.9%, and the mating efficiency between *S. cerevisiae* strains was 19%. Hence, intraspecific mating efficiency was 21 times higher than interspecific mating.

We next assayed the entire deletion collection for mating efficiency with *S. uvarum* in order to find strains that either increased or decreased mating. These mating assays showed consistent effects between biological replicates (Supplemental Figure 1), and known sterile genes, as well as the lysine auxotrophy, demonstrate our ability to find sterile genes in the genome (Supplemental Table 1). We found that many strains had decreased mating relative to the average, and few with increased mating (Figure 1). Several new sterile genes were discovered using our technique – *RPB9, RPC25, YKL077W, SWC4, PRP28, SEC6, NSR1, POR1,* and *GRX5,* although no genes caused complete incompatibles only to interspecific mating. This is the first known description of the function of the putative gene *YKL077W*. *RPB9* and *RPC25* are members of RNA polymerase II and III complexes, respectively. *SWC4* alters chromatin and is a component of the NuA4 acetyltranferase complex (Krogan et al., 2003). *PRP28* is a ribosomal protein, and *NSR1* is involved in pre-rRNA processing. Together, these genes may be responsible for transcribing and translating crucial members of the mating pathway. *POR1* and *GRX5* are both nuclear proteins that localize to the mitochondria. Inheritance of mitochondria in yeast is considerably more complicated than in higher eukaryotes. Hence deleting these genes may cause mitochondrial incompatibilities in the mating pathway. Lastly, *SEC6* plays a role in cellular architecture and fuses secretory vesicles with the plasma membrane.

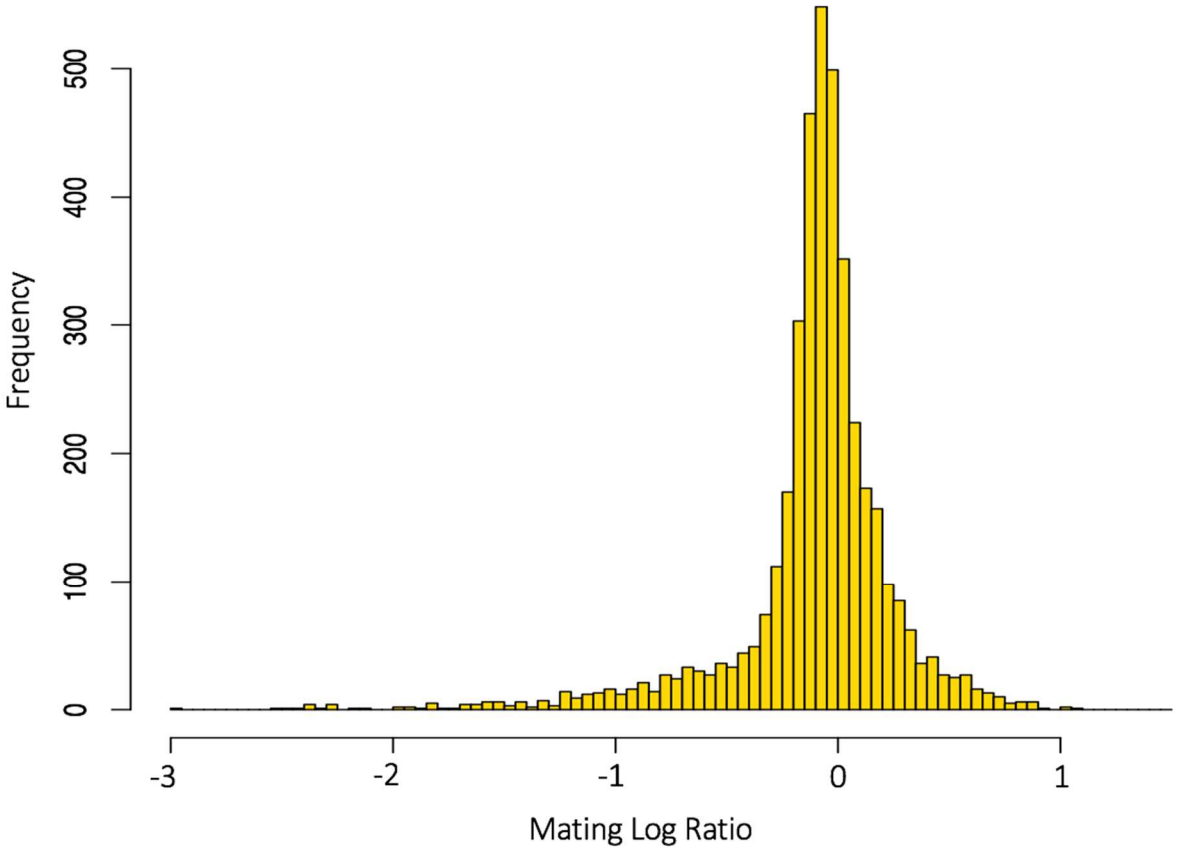

Figure 4.1. Distribution of interspecific pseudo-mating scores for all the nonessential genes in the deletion collection.

We individually assayed the mating efficiency of 41 deletion strains that represented the top and bottom of the mating distribution, and none had hybrid-specific mating incompatibilities. Several strains did have orders of magnitude drops in interspecific mating efficiency while maintaining near-normal intraspecific mating efficiency (Table 1). Both *ECM19* and *RRG8* are largely uncharacterized proteins that localize to the mitochondria and maintain the mitochondrial genome, respectively. No strains had significant increases in interspecific mating when individual mating assays were performed.

Table 4.6. Genes with low interspecific mating efficiency as determined from individual mating assays.

ORF	Gene Name	Interspecific Mating Efficiency	Intraspecific Mating Efficiency
YOR169C		0.00026	0.084
YER091C	*MET6*	0.00010	0.071
YLR390W	*ECM19*	0.00020	0.053
YDL236W	*PHO13*	0.00078	0.173
YPR116W	*RRG8*	0.00083	0.137

DISCUSSION

This is the first time, to our knowledge, that *S. cerevisiae* and *S. uvarum* mating efficiencies are directly compared, showing that interspecific mating is 21 fold less efficient than intraspecific mating. This mating preference would contribute to reproductive isolation between species.

After determining interspecific mating efficiencies for the majority of the deletion collection, we found no genes that are completely essential for interspecific mating but nonessential for intraspecific mating. These findings are consistent with other studies that have found no complete prezygotic barriers to speciation within the *Saccharomyces sensu stricto.* There were several deleted genes that caused decreased interspecific mating while maintaining normal intraspecific mating efficiency. Deleting these genes would decrease genetic flow between species, and may represent some of the potential small incompatibilities that add up to yield complete genetic isolation. Future studies should combine these mutations into a single

strain that would mate with *S. cerevisiae* but that has no or extremely low mating efficiency with *S. uvarum.*

The genes that decrease interspecific mating efficiency include two previously uncharacterized nuclear genes that locate to the mitochondria. Nuclear-mitochondrial incompatibilities have been shown to cause sterility in hybrid *Saccharomyces,* and since the mitochondrial fusion in yeast is particularly complicated during mating, incompatibilities in this pathway may partially prevent interspecific mating. This may be the first observation of hybrid nuclear-mitochondrial incompatibilities in the mating pathway.

This study also discovered several new sterile genes, and also found potential function for the putative gene *YKL077W.* To our knowledge, this is the first described function of this gene. Several of these sterile genes are involved in transcription and translation, and may be responsible for expressing key members of the mating pathway. Furthermore, 1/3 of the genome, including the essential genes, are yet to be assayed for their contributions to interspecific mating.

Overall, our results show that interspecific mating in the *sensu stricto* is robust, though not uniform between strains. The presence of multiple strains that hinder only interspecific mating by over an order of magnitude suggests that it is possible to create a *S. cerevisiae* strain that cannot mate *S. uvarum*. Such a strain would be the first example of complete reproductive isolation between the two species.

SUPPLEMENT

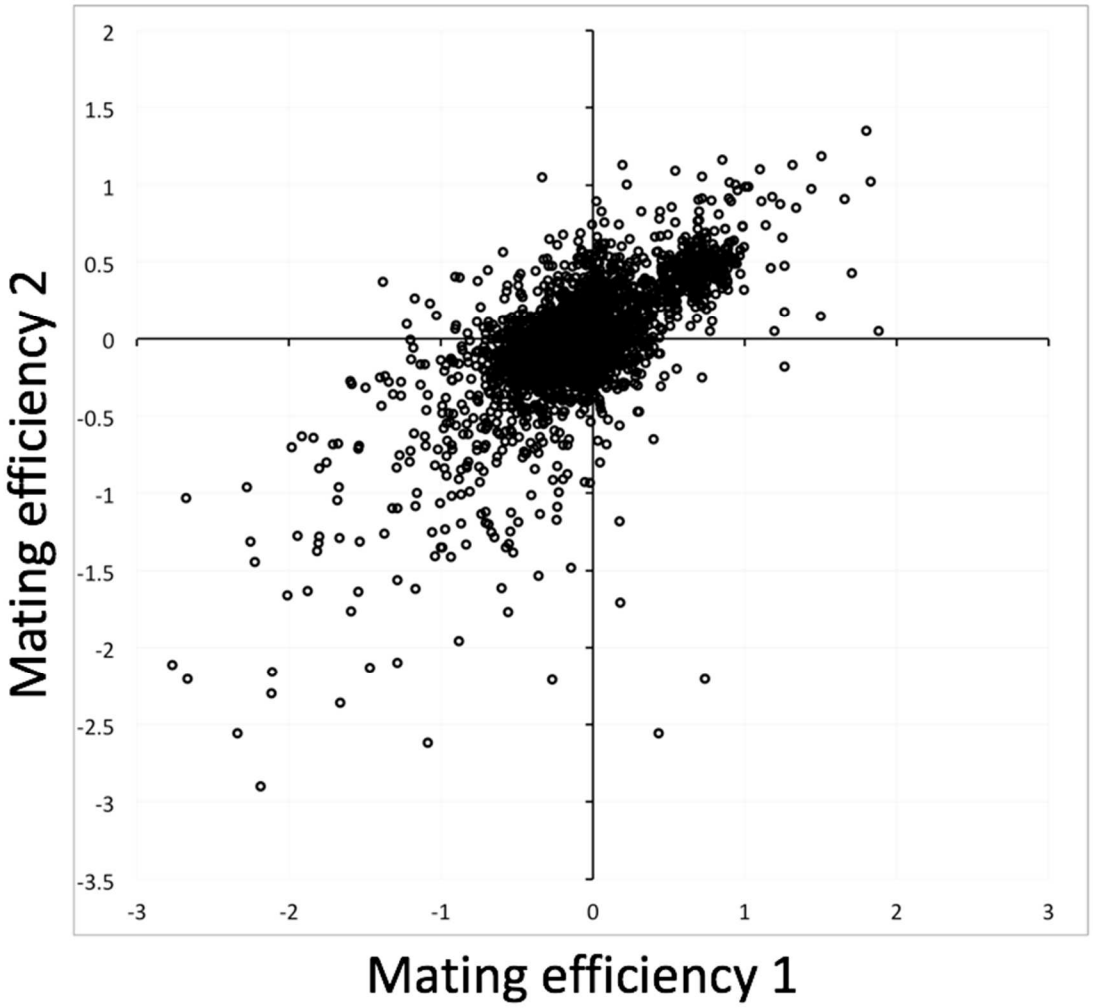

Figure 4.2. Biological replicates of the interspecific deletion collection matings. Most strains have mating effects that are consistent between biological replicates. R^2=0.5

Table 4.7. Sterile genes present in the data. Reads normalized to total reads in the sequencing run.

Gene	**Total Reads Before Mating**	**Normalized Reads**	**Total Reads After Mating**	**Normalized Reads**	**Log Ratio**
ste2	405	0.00021	1	0.00000	-2.67
ste20	683	0.00036	6	0.00000	-2.12
ste6	528	0.00028	13	0.00001	-1.67
ste24	968	0.00051	0	0.00000	
ste4	80	0.00004	1	0.00000	-1.96
ste11	207	0.00011	1	0.00000	-2.38
ste23	430	0.00023	273	0.00013	-0.26
ste3	405	0.00021	244	0.00011	-0.28
lys2	61	0.00003	0	0.00000	

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

GENOMICS AND PROTEOMICS IN HYBRIDS

The impact of this work extends to several fields. It has been shown previously that hemizygous gene deletions in diploid *S. cerevisiae* have environment-specific effects, and I have shown that this true in hybrids. I went on to show that the effects of hemizygous deletions are dependent on genetic background too. Simply put, deleting an allele in a purebred in many instances will not predict the effects of that deletion in the hybrid. These findings have implications any time DNA is introduced into a new genetic background, as any horizontally transferred genes may act very differently in a new organism. Further, knowing the effects of an intervention in one organism might not predict its effects in another. These ideas highlight the importance of experiments because predicting the function of genes, even with sophisticated computational techniques, may be extremely difficult.

We do, however, see more consistency between environments when GO terms are examined than for individual genes. Further, our GO findings for the purebreds overlapped with what other laboratories have found. This suggests that certain pathways are consistently important in a given genetic background. Future studies would identify whole pathways or genetic networks that contribute to phenotypes in hybrids, and determine if disrupting that pathway would have conserved effects across genetic backgrounds. Perhaps this could be accomplished by inducibly knocking out portions of genetic networks in the purebreds and hybrids. Furthermore, looking for enrichments in the data using the genetic network from the synthetic gene array (Costanzo et al., 2010) would provide a better view of networks than GO.

The gene by background interactions highlight an interesting fact – that yeast hybrids have different genetic means for moving around a fitness landscape than the purebreds. This may

explain why hybrids often have such unique karyotypes, because once formed, their genome may be altered in new ways to explore peaks on the fitness landscape. This contrasts with purebreds whose genomes are more closely optimized, and move around the fitness landscape in a more constrained way. Related to this idea, some cancers have genomes dramatically altered by mutagenic oncogenes. Once substantially altered, their genomes are no longer nearly optimized and have new genetic routes through fitness space. Longer term batch evolution of hybrids might reveal more genome flexibility than has been seen so far in the lab.

Additional future studies should further elucidate the causes of heterosis in the hybrid deletion collection. In particular, the large peak of high fitness strains in phosphate limited media remains unexplained, and taking two clones from this population and determining a biochemical and molecular explanation for the fitness effects observed would be invaluable. Do these clones adapt to the phosphate limitation in a similar way, or are they entirely distinct from one another? If there is an incompatible genetic network represented by this peak, it might be a Dobzhansky-Muller-like incompatibility at the network level.

The method I helped pioneer for finding the strength of interactions between different proteins confirms that highly diverged proteins interact in hybrids. Such knowledge suggests that most changes in amino acid sequence do not change the fundamental structure of the proteins. Despite the divergence within the *Saccharomyces sensu stricto*, it seems protein-protein interactions are robust. As such, determining which residues are necessary for interactions would be an important future step to understanding how heterogeneous protein complexes form. Furthermore, determining at which threshold of expression the composition of these complexes changes is an important step in linking expression with composition. This could be done by

titrating mRNA levels of a particular gene with a known preference score to see when that score changes.

We can reconstruct the average protein complex in the hybrid, showing that in most cases the complexes are heterogeneous. We have created a model of the average hybrid SAGA complex, which recruits transcription factors, and have shown that some orthologs are consistently preferred over others. If such heterogeneous complexes contribute to expression of genetic networks in the hybrid, they may be responsible for the dominant network wiring that confers heterosis. Linking up preference scores with SAGA-related expression may show how genetic networks are regulated in the hybrids. This could be accomplished by deleting members of the complex, determining its composition, and then finding how the deletion affects gene expression at the genome level.

PREZYGOTIC BARRIERS IN THE SENSU STRICTO

I also investigated prezygotic barriers to hybridization between *S. uvarum* and *S. cerevisiae.* My screen of over 4,000 strains showed that no deleted genes completely prevented mating between the species, although there were some that dramatically decreased interspecific mating. This study is consistent with other studies in the *Saccharomyces sensu stricto* that have shown interspecific mating is robust, but that suggest many small contributions to reproductive isolation could potentially add up to create complete isolation. Perhaps complete isolation could be accomplished by combining all the deletions that reduce interspecific mating into a single strain. Understanding reproductive isolation in the *Saccharomyces sensu stricto* remains elusive,

although experiments in this dissertation have shown that creating reproductively isolated strains may be possible.

CHAPTER 6: CITATIONS

Antunovics, Z., Nguyen, H.-V., Gaillardin, C., & Sipiczki, M. (2005). Gradual genome stabilisation by progressive reduction of the Saccharomyces uvarum genome in an interspecific hybrid with Saccharomyces cerevisiae. *FEMS Yeast Research*, *5*(12), 1141–1150. http://doi.org/10.1016/j.femsyr.2005.04.008

Artieri, C. G., & Fraser, H. B. (2014). Evolution at two levels of gene expression in yeast. *Genome Research*, *24*(3), 411–421. http://doi.org/10.1101/gr.165522.113

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics (Oxford, England)*, *25*(8), 1091–1093. http://doi.org/10.1093/bioinformatics/btp101

Blein-Nicolas, M., Albertin, W., da Silva, T., Valot, B., Balliau, T., Masneuf-Pomarède, I., et al. (2015). A systems approach to elucidate heterosis of protein abundances in yeast. *Molecular & Cellular Proteomics*, *14*(8), mcp.M115.048058–2071. http://doi.org/10.1074/mcp.M115.048058

Brem, R. B., Yvert, G., Clinton, R., & Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science*, *296*(5568), 752–755. http://doi.org/10.1126/science.1069516

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L. A., D'Amore, R., Allen, A. M., et al. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, *491*(7426), 705. http://doi.org/10.1038/nature11650

CC, L., JE, B., & JD, D. C. (1953). Heterosis in Saccharomyces. *Nature*, *172*(4383), 800–802.

Chambers, S. R., Hunter, N., Louis, E. J., & Borts, R. H. (1996). The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Molecular and Cellular Biology*, *16*(11), 6110–6120. http://doi.org/10.1128/MCB.16.11.6110

Chou, J.-Y., Hung, Y.-S., Lin, K.-H., Lee, H.-Y., & Leu, J.-Y. (2010). Multiple Molecular Mechanisms Cause Reproductive Isolation between Three Yeast Species. *PLOS Biol*, *8*(7), e1000432. http://doi.org/10.1371/journal.pbio.1000432

Connelly, C. F., Wakefield, J., & Akey, J. M. (2014). Evolution and Genetic Architecture of Chromatin Accessibility and Function in Yeast. *PLoS Genet*, *10*(7), e1004427. http://doi.org/10.1371/journal.pgen.1004427

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., et al. (2010). The Genetic Landscape of a Cell. *Science*, *327*(5964), 425–431. http://doi.org/10.1126/science.1180823

Csoma, H., Zakany, N., Capece, A., Romano, P., & Sipiczki, M. (2010). Biological diversity of Saccharomyces yeasts of spontaneously fermenting wines in four wine regions: comparative genotypic and phenotypic analysis. *International Journal of Food Microbiology*, *140*(2-3), 239–248. http://doi.org/10.1016/j.ijfoodmicro.2010.03.024

Darwin, C. (1909). The Origin of Species. Dent.

de Barros Lopes, M., Bellon, J. R., Shirley, N. J., & Ganter, P. F. (2002). Evidence for multiple interspecific hybridization in Saccharomyces sensu stricto species. *FEMS*

Yeast Research, *1*(4), 323–331.
Delneri, D., Hoyle, D. C., Gkargkas, K., Cross, E. J. M., Rash, B., Zeef, L., et al. (2008). Identification and characterization of high-flux-control genes of yeast through competition analyses in continuous cultures. *Nature Genetics*, *40*(1), 113–117. http://doi.org/10.1038/ng.2007.49
Deutschbauer, A. M., & Davis, R. W. (2005). Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nature Genetics*, *37*(12), 1333–1340. http://doi.org/10.1038/ng1674
DEVRIES, P. J., AUSTIN, G. T., & MARTIN, N. H. (2008). Diel activity and reproductive isolation in a diverse assemblage of Neotropical skippers (Lepidoptera: Hesperiidae). *Biological Journal of the Linnean Society*, *94*(4), 723–736. http://doi.org/10.1111/j.1095-8312.2008.01037.x
Dietrich, F. S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., et al. (2004). The Ashbya gossypii Genome as a Tool for Mapping the Ancient Saccharomyces cerevisiae Genome. *Science*, *304*(5668), 304–307. http://doi.org/10.1126/science.1095781
Dunn, B., & Sherlock, G. (2008). Reconstruction of the genome origins and evolution of the hybrid lager yeast Saccharomyces pastorianus. *Genome Research*, *18*(10), 1610–1623. http://doi.org/10.1101/gr.076075.108
Dunn, B., Paulish, T., Stanbery, A., Piotrowski, J., Koniges, G., Kroll, E., et al. (2013). Recurrent Rearrangement during Adaptive Evolution in an Interspecific Yeast Hybrid Suggests a Model for Rapid Introgression. *PLoS Genet*, *9*(3), e1003366. http://doi.org/10.1371/journal.pgen.1003366
EC, S., 3rd, Y. J., & L, G. (2003). MudPIT: A powerful proteomics tool for discovery. *Discovery Medicine*, *3*(18), 38–39.
Ehrenreich, I. M., Gerke, J. P., & Kruglyak, L. (2009). Genetic Dissection of Complex Traits in Yeast: Insights from Studies of Gene Expression and Other Phenotypes in the BY×RM Cross. *Cold Spring Harbor Symposia on Quantitative Biology*, *74*(0), 145–153. http://doi.org/10.1101/sqb.2009.74.013
Ellis, A. G., Weis, A. E., & Gaut, B. S. (2006). EVOLUTIONARY RADIATION OF “STONE PLANTS” IN THE GENUS ARGYRODERMA (AIZOACEAE): UNRAVELING THE EFFECTS OF LANDSCAPE, HABITAT, AND FLOWERING TIME. *Evolution*, *60*(1), 39–55. http://doi.org/10.1111/j.0014-3820.2006.tb01080.x
Fischer, G., James, S. A., Roberts, I. N., Oliver, S. G., & Louis, E. J. (2000). Chromosomal evolution in Saccharomyces. *Nature*, *405*(6785), 451–454. http://doi.org/10.1038/35013058
Fleck, O., Lehmann, E., Schär, P., & Kohli, J. (1999). Involvement of nucleotide-excision repair in msh2 pms1-independent mismatch repair. *Nature Genetics*, *21*(3), 314–317. http://doi.org/10.1038/6838
Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., et al. (2003). Global analysis of protein expression in yeast. *Nature*, *425*(6959), 737–741. http://doi.org/10.1038/nature02046
Giaever, G., & Nislow, C. (2014). The Yeast Deletion Collection: A Decade of Functional Genomics. *Genetics*, *197*(2), 451–465. http://doi.org/10.1534/genetics.114.161620
Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, *418*(6896),

387–391. http://doi.org/10.1038/nature00935
Goddard, M. R., & Greig, D. (2015). Saccharomyces cerevisiae: a nomadic yeast with no niche? *FEMS Yeast Research*.
Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 Genes. *Science*, *274*(5287), 546–567. http://doi.org/10.1126/science.274.5287.546
González, S. S., Barrio, E., Gafner, J., & Querol, A. (2006). *FEMS Yeast Research*, (8). http://doi.org/10.1111/j.1567-1364.2006.00126.x
Gordon, D. J., Resio, B., & Pellman, D. (2012). Causes and consequences of aneuploidy in cancer. *Nature Reviews Genetics*, *13*(3), 189–203. http://doi.org/10.1038/nrg3123
Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. A., Ward, A., et al. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet*, *4*(12), e1000303. http://doi.org/10.1371/journal.pgen.1000303
Han, Y., Luo, J., Ranish, J., & Hahn, S. (2014). Architecture of the Saccharomyces cerevisiae SAGA transcription coactivator complex. *The EMBO Journal*, *33*(21), 2534–2546. http://doi.org/10.15252/embj.201488638
Hayden, B., & Villeneuve, S. (2011). A Century of Feasting Studies. *Dx.Doi.org*, *40*(1), 433–449. http://doi.org/10.1146/annurev-anthro-081309-145740
Hittinger, C. T. (2013). Saccharomyces diversity and evolution: a budding model genus. *Trends in Genetics*, *29*(5), 309–317. http://doi.org/10.1016/j.tig.2013.01.002
Ho, C. H., Magtanong, L., Barker, S. L., Gresham, D., Nishimura, S., Natarajan, P., et al. (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nature Biotechnology*, *27*(4), 369–377. http://doi.org/10.1038/nbt.1534
Hodgkin, J., Jackson, I., Mak, Y., Martienssen, R., & Rowley, A. (n.d.). Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Cell*. http://doi.org/10.1016/pic.2015.09.24.2351
Horinouchi, T., Yoshikawa, K., Kawaide, R., Furusawa, C., Nakao, Y., Hirasawa, T., & Shimizu, H. (2010). Genome-wide expression analysis of Saccharomyces pastorianus orthologous genes using oligonucleotide microarrays. *Journal of Bioscience and Bioengineering*, *110*(5), 602–607. http://doi.org/10.1016/j.jbiosc.2010.05.011
Hunter, N., Chambers, S. R., Louis, E. J., & Borts, R. H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *The EMBO Journal*, *15*(7), 1726–1733.
IWGSC, T. I. W. G. S. C., Mayer, K. F. X., Rogers, J., Doležel, J., Pozniak, C., Eversole, K., et al. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*, *345*(6194), 1251788–1251788. http://doi.org/10.1126/science.1251788
Johmura, Y., Soung, N.-K., Park, J.-E., Yu, L.-R., Zhou, M., Bang, J. K., et al. (2011). Regulation of microtubule-based microtubule nucleation by mammalian polo-like kinase 1. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(28), 11446–11451. http://doi.org/10.1073/pnas.1106223108
Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of

ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, *428*(6983), 617–624. http://doi.org/10.1038/nature02424
Kellis, M., Patterson, N., Endrizzi, M., Birren, B., & Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, *423*(6937), 241–254. http://doi.org/10.1038/nature01644
Khan, Z., Bloom, J. S., Amini, S., Singh, M., Perlman, D. H., Caudy, A. A., & Kruglyak, L. (2012). Quantitative measurement of allele-specific protein expression in a diploid yeast hybrid by LC-MS. *Molecular Systems Biology*, *8*, 602. http://doi.org/10.1038/msb.2012.34
Krogan, N. J., Keogh, M.-C., Datta, N., Sawa, C., Ryan, O. W., Ding, H., et al. (2003). A Snf2 Family ATPase Complex Required for Recruitment of the Histone H2A Variant Htz1. *Molecular Cell*, *12*(6), 1565–1576. http://doi.org/10.1016/S1097-2765(03)00497-0
Kunicka-Styczyńska, A., & Rajkowska, K. (2011). Physiological and genetic stability of hybrids of industrial wine yeasts Saccharomyces sensu stricto complex. *Journal of Applied Microbiology*, *110*(6), 1538–1549. http://doi.org/10.1111/j.1365-2672.2011.05009.x
Leducq, J.-B., Charron, G., Diss, G., Gagnon-Arsenault, I., Dubé, A. K., & Landry, C. R. (2012). Evidence for the Robustness of Protein Complexes to Inter-Species Hybridization. *PLoS Genet*, *8*(12), e1003161. http://doi.org/10.1371/journal.pgen.1003161
Lee, H.-Y., Chou, J.-Y., Cheong, L., Chang, N.-H., Yang, S.-Y., & Leu, J.-Y. (2008). Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility between Two Yeast Species. *Cell*, *135*(6), 1065–1073. http://doi.org/10.1016/j.cell.2008.10.047
Leu, J.-Y., & Murray, A. W. (2006). Experimental Evolution of Mating Discrimination in Budding Yeast. *Current Biology*, *16*(3), 280–286. http://doi.org/10.1016/j.cub.2005.12.028
Li, Z., Vizeacoumar, F. J., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F. S., et al. (2011). Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nature Biotechnology*, *29*, 367. http://doi.org/10.1038/nbt.1832
Libkind, D., Hittinger, C. T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., et al. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proceedings of the National Academy of Sciences*, *108*(35), 14539–14544. http://doi.org/10.1073/pnas.1105430108
Lin, T.-C., Gombos, L., Neuner, A., Sebastian, D., Olsen, J. V., Hrle, A., et al. (2011). Phosphorylation of the Yeast γ-Tubulin Tub4 Regulates Microtubule Function. *PLoS One*, *6*(5), e19700. http://doi.org/10.1371/journal.pone.0019700
Liti, G. (2015). The fascinating and secret wild life of the budding yeast S. cerevisiae. *eLife*, *4*, e05835. http://doi.org/10.7554/eLife.05835
Liti, G., & Louis, E. J. (2005). Yeast evolution and comparative genomics. *Annual Review of Microbiology*, *59*, 135–153. http://doi.org/10.1146/annurev.micro.59.030804.121400
Maclean, C. J., & Greig, D. (2008). Prezygotic reproductive isolation between Saccharomyces cerevisiae and Saccharomyces paradoxus. *BMC Evolutionary Biology*, *8*(1), 1. http://doi.org/10.1186/1471-2148-8-1

Marcet-Houben, M., & Gabaldón, T. (2015). Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker's yeast lineage. *PLOS Biol*.

Marcireau, C., Joets, J., Pousset, D., Guilloton, M., & Karst, F. (1996). FEN2: a gene implicated in the catabolite repression-mediated regulation of ergosterol biosynthesis in yeast. *Yeast*, *12*(6), 531–539. http://doi.org/10.1002/(SICI)1097-0061(199605)12:6<531::AID-YEA934>3.0.CO;2-E

Martin, O. C., DeSevo, C. G., Guo, B. Z., Koshland, D. E., Dunham, M. J., & Zheng, Y. (2009). Telomere behavior in a hybrid yeast. *Cell Research*, *19*(7), 910–912. http://doi.org/10.1038/cr.2009.65

Martini, A. V., & Martini, A. (1987). Three newly delimited species of Saccharomyces sensu stricto. *Antonie Van Leeuwenhoek*, *53*(2), 77–84.

McElfresh, J. S., & Millar, J. G. (2008). GEOGRAPHIC VARIATION IN THE PHEROMONE SYSTEM OF THE SATURNIID MOTHHEMILEUCA EGLANTERINA. *Dx.Doi.org*, *82*(12), 3505–3518. http://doi.org/10.1890/0012-9658(2001)082[3505:GVITPS]2.0.CO;2

McGee, H. (2007). On food and cooking: the science and lore of the kitchen.

Mellacheruvu, D., Wright, Z., Couzens, A. L., Lambert, J.-P., St-Denis, N. A., Li, T., et al. (2013). The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature Methods*, *10*(8), 730–736. http://doi.org/10.1038/nmeth.2557

Merlini, L., Dudin, O., & Martin, S. G. (2013). Mate and fuse: how yeast cells do it. *Open Biology*, *3*(3), 130008–130008. http://doi.org/10.1098/rsob.130008

Mnaimneh, S., Davierwala, A. P., Haynes, J., Moffat, J., Peng, W.-T., Zhang, W., et al. (2004). Exploration of essential gene functions via titratable promoter alleles. *Cell*, *118*(1), 31–44. http://doi.org/10.1016/j.cell.2004.06.013

Muller, L. A. H., & McCusker, J. H. (2009). A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in Saccharomyces cerevisiae. *FEMS Yeast Research*, *9*(1), 143–152. http://doi.org/10.1111/j.1567-1364.2008.00464.x

Murphy, H. A., & Zeyl, C. W. (2012). PREZYGOTIC ISOLATION BETWEEN SACCHAROMYCES CEREVISIAE AND SACCHAROMYCES PARADOXUS THROUGH DIFFERENCES IN MATING SPEED AND GERMINATION TIMING. *Evolution*, *66*(4), 1196–1209. http://doi.org/10.1111/j.1558-5646.2011.01516.x

Nakao, Y., Kanamori, T., Itoh, T., Kodama, Y., Rainieri, S., Nakamura, N., et al. (2009). Genome Sequence of the Lager Brewing Yeast, an Interspecies Hybrid. *DNA Research*, *16*(2), 115–129. http://doi.org/10.1093/dnares/dsp003

Naumov, G., Naumova, E., & Gaillardin, C. (1993). Genetic and Karyotypic Identification of Wine Saccharomyces bayanus Yeasts Isolated in France and Italy. *Systematic and Applied Microbiology*, *16*(2), 274–279. http://doi.org/10.1016/S0723-2020(11)80480-7

Naumova, E. S., Naumov, G. I., Masneuf-Pomarède, I., Aigle, M., & Dubourdieu, D. (2005). Molecular genetic study of introgression between Saccharomyces bayanus and S. cerevisiae. *Yeast*, *22*(14), 1099–1115. http://doi.org/10.1002/yea.1298

Nguyen, H.-V., Legras, J.-L., Neuvéglise, C., & Gaillardin, C. (2011). Deciphering the Hybridisation History Leading to the Lager Lineage Based on the Mosaic Genomes

of Saccharomyces bayanus Strains NBRC1948 and CBS380 T. *PLoS One*, *6*(10), e25821. http://doi.org/10.1371/journal.pone.0025821

Nguyen, H.-V., Lepingle, A., & Gaillardin, C. (2000). Molecular Typing Demonstrates Homogeneity of Saccharomyces uvarum Strains and Reveals the Existence of Hybrids between S. uvarum and S. cerevisiae, Including the S. bayanus Type Strain CBS 380. *Systematic and Applied Microbiology*, *23*(1), 71–85. http://doi.org/10.1016/S0723-2020(00)80048-X

Novo, M., Mangado, A., Quirós, M., Morales, P., Salvadó, Z., & Gonzalez, R. (2013). Genome-wide study of the adaptation of Saccharomyces cerevisiae to the early stages of wine fermentation. *PLoS One*, *8*(9), e74086. http://doi.org/10.1371/journal.pone.0074086

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., & Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics*, *1*(5), 376–386. http://doi.org/10.1074/mcp.M200025-MCP200

Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics*, *34*, 401–437. http://doi.org/10.1146/annurev.genet.34.1.401

Park, S. K. R., Aslanian, A., McClatchy, D. B., Han, X., Shah, H., Singh, M., et al. (2014). Census 2: isobaric labeling data analysis. *Bioinformatics (Oxford, England)*, *30*(15), btu151–2209. http://doi.org/10.1093/bioinformatics/btu151

Pellman, D. (2007). Cell biology: aneuploidy and cancer. *Nature*.

Peris, D., Lopes, C. A., Belloch, C., Querol, A., & Barrio, E. (2012). Comparative genomics among Saccharomyces cerevisiae × Saccharomyces kudriavzevii natural hybrid strains isolated from wine and beer reveals different origins. *BMC Genomics*, *13*(1), 407. http://doi.org/10.1186/1471-2164-13-407

Pérez-Través, L., Lopes, C. A., Querol, A., & Barrio, E. (2014). On the Complexity of the Saccharomyces bayanus Taxon: Hybridization and Potential Hybrid Speciation. *PLoS One*, *9*(4), e93729. http://doi.org/10.1371/journal.pone.0093729

Pfliegler, W. P., Antunovics, Z., & Sipiczki, M. (2012). Double sterility barrier between Saccharomyces species and its breakdown in allopolyploid hybrids by chromosome loss. *FEMS Yeast Research*, *12*(6), 703–718. http://doi.org/10.1111/j.1567-1364.2012.00820.x

Phipps, K. R., Charette, J. M., & Baserga, S. J. (2011). The small subunit processome in ribosome biogenesis—progress and prospects. *Wiley Interdisciplinary Reviews: RNA*, *2*(1), 1–21. http://doi.org/10.1002/wrna.57

Piatkowska, E. M., Naseeb, S., Knight, D., & Delneri, D. (2013). Chimeric Protein Complexes in Hybrid Species Generate Novel Phenotypes. *PLoS Genet*, *9*(10), e1003836. http://doi.org/10.1371/journal.pgen.1003836

Piotrowski, J. S., Nagarajan, S., Kroll, E., Stanbery, A., Chiotti, K. E., Kruckeberg, A. L., et al. (2012). Different selective pressures lead to different genomic outcomes as newly-formed hybrid yeasts evolve. *BMC Evolutionary Biology*, *12*, 46. http://doi.org/10.1186/1471-2148-12-46

Ramani, A. K., & Marcotte, E. M. (2003). Exploiting the Co-evolution of Interacting Proteins to Discover Interaction Specificity. *Journal of Molecular Biology*, *327*(1), 273–284. http://doi.org/10.1016/S0022-2836(03)00114-1

Rappsilber, J., Mann, M., & Ishihama, Y. (2007). Protocol for micro-purification,

enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols*, *2*(8), 1896–1906. http://doi.org/10.1038/nprot.2007.261

Rogers, D. W., Denton, J. A., McConnell, E., & Greig, D. (2015). Experimental Evolution of Species Recognition. *Current Biology*, *25*(13), 1753–1758. http://doi.org/10.1016/j.cub.2015.05.023

Ruderfer, D. M., Pratt, S. C., Seidel, H. S., & Kruglyak, L. (2006). Population genomic analysis of outcrossing and recombination in yeast. *Nature Genetics*, *38*(9), 1077–1081. http://doi.org/10.1038/ng1859

Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S., & Wolfe, K. H. (2006). Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature*, *440*(7082), 341–345. http://doi.org/10.1038/nature04562

Scannell, D. R., Frank, A. C., Conant, G. C., Byrne, K. P., Woolfit, M., & Wolfe, K. H. (2007). Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proceedings of the National Academy of Sciences*, *104*(20), 8397–8402. http://doi.org/10.1073/pnas.0608218104

Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., et al. (2011). The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. *G3: Genes, Genomes, Genetics*, *1*(1), 11–25. http://doi.org/10.1534/g3.111.000273

Sebastiani, F., Barberio, C., Casalone, E., Cavalieri, D., & Polsinelli, M. (2002). Crosses between Saccharomyces cerevisiae and Saccharomyces bayanus generate fertile hybrids. *Research in Microbiology*, *153*(1), 53–58. http://doi.org/10.1016/S0923-2508(01)01286-4

Sia, E. A., Kokoska, R. J., Dominska, M., Greenwell, P., & Petes, T. D. (1997). Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Molecular and Cellular Biology*, *17*(5), 2851–2858.

Sipiczki, M. (2008). Interspecies hybridization and recombination in Saccharomyces wine yeasts. *FEMS Yeast Research*, *8*(7), 996–1007. http://doi.org/10.1111/j.1567-1364.2008.00369.x

Smith, E. N., & Kruglyak, L. (2008). Gene–Environment Interaction in Yeast Gene Expression. *PLOS Biol*, *6*(4), e83. http://doi.org/10.1371/journal.pbio.0060083

Smith, V., Botstein, D., & Brown, P. O. (1995). Genetic footprinting: a genomic strategy for determining a gene's function given its sequence. *Proceedings of the National Academy of Sciences*, *92*(14), 6479–6483.

Steinmetz, L. M., Sinha, H., Richards, D. R., Spiegelman, J. I., Oefner, P. J., McCusker, J. H., & Davis, R. W. (2002). Dissecting the architecture of a quantitative trait locus in yeast. *Nature*, *416*(6878), 326–330. http://doi.org/10.1038/416326a

Sterner, D. E., & Berger, S. L. (2000). Acetylation of Histones and Transcription-Related Factors. *Microbiology and Molecular Biology Reviews*, *64*(2), 435–459. http://doi.org/10.1128/MMBR.64.2.435-459.2000

Stolz, J., & Sauer, N. (1999). The fenpropimorph resistance gene FEN2 from Saccharomyces cerevisiae encodes a plasma membrane H+-pantothenate symporter. *The Journal of Biological Chemistry*, *274*(26), 18747.

Tannahill, R. (1989). Food in history. Grown trade paperbacks.

Tirosh, I., Reikhav, S., Levy, A. A., & Barkai, N. (2009). A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science*, *324*(5927), 659–662. http://doi.org/10.1126/science.1169766

Vogel, J., Drapkin, B., Oomen, J., Beach, D., Bloom, K., & Snyder, M. (2001). Phosphorylation of γ-Tubulin Regulates Microtubule Organization in Budding Yeast. *Developmental Cell*, *1*(5), 621–631. http://doi.org/10.1016/S1534-5807(01)00073-9

Wapinski, I., Pfeffer, A., Friedman, N., & Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. *Nature*, *449*(7158), 54–61. http://doi.org/10.1038/nature06107

Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., et al. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. *Science*, *285*(5429), 901–906. http://doi.org/10.1126/science.285.5429.901

Wolfe, K. H., & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*.

Woolford, J. L., & Baserga, S. J. (2013). Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics*, *195*(3), 643–681. http://doi.org/10.1534/genetics.113.153197

Yan, Z., Costanzo, M., Heisler, L. E., Paw, J., Kaper, F., Andrews, B. J., et al. (2008). Yeast Barcoders: a chemogenomic application of a universal donor-strain collection carrying bar-code identifiers. *Nature Methods*, *5*(8), 719–725. http://doi.org/10.1038/nmeth.1231

Yoon, O. K., & Brem, R. B. (2010). Noncanonical transcript forms in yeast and their regulation during environmental stress. *RNA (New York, N.Y.)*, *16*(6), 1256–1267. http://doi.org/10.1261/rna.2038810