Genetics of adaptation in experimental populations of yeast

by

Jasmine Ono

B.Sc., The University of Toronto, 2008
M.Sc., The University of British Columbia, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2017

Abstract

Evolution proceeds through genetic changes to individuals, which are either propagated or disappear over generations. Adaptation is one of the main mechanisms driving these changes in genetic composition. Speciation can also result from different, and incompatible, genetic changes occurring in different populations. This thesis furthers our knowledge of the genetics of adaptation and speciation using the budding yeast *Saccharomyces cerevisiae*. My work on the genetic basis of adaptation to high concentrations of copper, when contrasted with a similar experiment using the fungicide nystatin, showed that the environment has a strong influence on both the number of genes that are the targets of selection and the types of potentially beneficial mutations. These results have implications for the repeatability of genetic evolution. In a second study, I found that genetic interactions between individually isolated single-step beneficial mutations from the same selective environment often exhibited the type of epistasis that underlies speciation even though these mutations occurred within a single biosynthetic pathway. These results support the mutation-order model of speciation by adaptation, where the chance order of mutations in separated populations leads to divergence and the build-up of reproductive isolation due to genetic incompatibility. Negative genetic interactions became positive when the level of stress was increased, indicating that genetically-based reproductive isolation can also be environment-dependent. Finally, I found that diploid yeast were generally not able to adapt to a level of fungicide to which haploid yeast can adapt. Diploids have been found to adapt to a lower concentration of the same drug, indicating that the exact environment (type and concentration) and ploidy can have an impact on the likelihood of genetic rescue. Together, these results have implications for our understanding of the genetic basis of adaptation in different types of environments and different levels of the same environmental stressor.

Lay Summary

Many aspects of an organism are encoded in their DNA, including their ability to tolerate stressful environments. Changes to DNA can therefore change how well suited an organism is to certain conditions, and better-adapted types have an advantage over others. Throughout this thesis I have studied such changes using experimentally-evolved populations of the common brewing and baking yeast, *Saccharomyces cerevisiae*. In doing so, I have found that the process of adaptation, and the underlying changes involved, can depend on a variety of factors including the nature of the environment, the level of stress imposed and the genome of the organism. In addition, I find that different genetic solutions to the same adaptive problem are not always compatible with each other. This incompatibility can lead to the evolution of new species.

Preface

A version of Chapter 2 has been published as "Gerstein, A. C., Ono, J., Lo, D. S., Campbell, M. L., Kuzmin, A., & Otto, S. P. (2015). Too much of a good thing: the unique and repeated paths toward copper adaptation. Genetics, 199(2), 555-571" with A.C. Gerstein and I listed as co-first authors. A.C. Gerstein and S.P. Otto conceived the original project, and I conceived and carried out the genetic analyses and subsequent assays of the lines to determine which mutations caused adaptation. I performed laboratory work in conjunction with A.C. Gerstein, D.S. Lo, M.L. Campbell and A. Kuzmin. S.P. Otto wrote the scripts to analyze the Illumina sequence data and determined the expected frequency of mutations causing nonsynonymous and stop codons. I performed analyses of copper tolerance of deletion lines and of single mutations and A.C. Gerstein performed other phenotypic analyses, with advice from S.P. Otto. The manuscript was written in collaboration with A.C. Gerstein and S.P. Otto. I wrote the initial draft of the sections on single mutations and the discussion and A.C. Gerstein, S.P. Otto and I all contributed major revisions to the manuscript.

A version of Chapter 3 has been published as "Ono, J., Gerstein, A. C., & Otto, S. P. (2017). Widespread genetic incompatibilities between first-step mutations during parallel adaptation of *Saccharomyces cerevisiae* to a common environment. PLoS biology, 15(1), e1002591". I conceived the project in conjunction with A.C. Gerstein and S.P. Otto. I performed the laboratory work and performed the analyses with advice from A.C. Gerstein and S.P. Otto. I prepared the majority of the visualizations, although Fig B.4 was prepared by A.C. Gerstein. I prepared the majority of the initial draft, with assistance from A.C. Gerstein and S.P. Otto. We all contributed major revisions to the manuscript.

A version of Chapter 4 is in preparation for publication in collaboration with A. Kuzmin, L. Miller and S.P. Otto. S.P. Otto and I conceived the original project. I performed laboratory work in conjunction with A. Kuzmin and L. Miller. S.P. Otto wrote the code to determine the expected number of mutations per site. I performed all analyses and prepared the initial draft of the manuscript, and S.P. Otto contributed revisions to the manuscript.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I have to start by thanking my supervisor, Sally Otto. I was very lucky to be in your lab, and I can only hope that others have the opportunity to have a supervisor like you. You have pushed me when I needed pushing and have been thoughtful and compassionate when I needed understanding. On top of that, I couldn't have asked for a better role model for how to be a scientist. Thank you for helping me to grow, both as an academic and as a person.

I would also like to thank Aleeza Gerstein. I have been lucky to have you as a mentor throughout my PhD. You are a co-author on two of my chapters and a spiritual co-author on the third, and collaborating with you has been one of the great joys of my PhD. Also, you give the best pep talks. Yeast friends 4 ever.

Thank you to my committee, Dolph Schluter, Michael Doebeli and Corey Nislow. Your advice and input have helped to expand my thinking and improve my science.

None of this would have been possible without all of the amazing people who've worked in the wet lab over the course of my PhD. Thank you to Dara Lo, William Li, Angelica Lillico-Ouachour, Andy An, Li-Wei Zhang, Ryan Stratychuk, and Chris James for their assistance in the lab, as well as Sherry Li and Anna Van Tol for their work before I arrived. A special thanks to Marcus Campbell for helping me to figure out how to supervise students and gain confidence in myself, Xinya Wang for bringing a steady and cheerful atmosphere to the lab, and to Lesley Miller for all of your help at the end of this long road. Of course, a big thank you to Anastasia Kuzmin, for all of your help but also for being there to talk to in the lab. Without you, I probably would have gone crazy long ago.

All of my lab mates have been great over the years, and I thank you all for your excitement and enthusiasm, whether it's for an epically long lab meeting or a ridiculous pumpkin carving. I especially want to thank Kate Ostevik, Nathaniel Sharp, Rich FitzJohn, Leithen M'Gonigle, Matt Osmond and Michael Scott.

Being the only yeast lab in our building, the Vancouver Yeast Group has been an amazing resource throughout my PhD, and many experiments would have been much more difficult without their help. A special thank you goes out to Krystina Ho, Vivien Measday and her lab, Chris Walkey and Allan DeBono for sharing their expertise and resources. Patricia Schulte and Tim Healy also provided expert assistance and advice when sorely needed.

A big thank you to everyone both in the Zoology Department and in the Biodiversity building who make grad school such a great place to be. Everyone who's ever organized a discussion group, a seminar speaker, a party or a symposium, or taken part in a skit, you are amazing. I especially want to thank some of the people who've made my own experience at UBC a special one including Kieran Samuk, Greg Owens, Alana Schick, Milica Mandic, Megan Bontrager, Kristin Nurkowski, Gina Conte, Brook Moyers, Kathryn Turner, Matt Siegle, Matt Pennell, Sam Yeaman, J.S. Moore, Anne Dalziel, Julie Lee-Yaw, and Kim Gilbert.

I also thank the funding sources that made these projects possible: the Four-Year fellowship from UBC

and funding from the National Science and Engineering Research Council of Canada (Discovery Grant to S.P. Otto).

Finally, a thank you to my family, to whom this thesis is dedicated, as well as to all of my 'unofficial' family: Courtenay Brown, Brett Willis, Vicki Lee, Nathaniel Brown, Kieran Bol, Grady Johnson and Ryan Vanderhoek. Thank you for still liking me even though I never email/text/call back. And, of course, to Dan Schick. You've been there for me through the whole thing. You've accepted my weird hours, messy office and disdain for dishes. It's hard to explain how much I appreciate you.

Dedication

This thesis is dedicated to my parents, George and Lulu, for their continued support, even if they didn't know what I was doing or why I was doing it.

Chapter 1

Introduction

We have come a long way since the time of Charles Darwin and Alfred Russel Wallace in our understanding of evolution (DARWIN 1859; DARWIN and WALLACE 1858). During the Modern Synthesis (roughly 1930s-1940s), ideas were developed about the interplay between genetics and evolution, and we have been testing and refining those ideas in ways that were unimaginable to those original scientists. Investigations of biological systems, both in the wild and in the lab, provide an expanding collection of case studies, which inform us about what has happened and how it has happened in specific instances. Evolutionary theory helps us to determine what patterns might arise and explores what might be possible, given a certain set of conditions. Experimental evolution borrows elements of both and can be found somewhere in between. In experimental evolution, we collect case studies of what actually happens given a certain set of conditions. Having control (although imperfect) over these conditions allows us to understand how slight differences in experimental input can lead to differences in biological output. Importantly, experimental evolution can tell us about the repeatability of such outcomes when the experiments are performed in replicate. Repeatability speaks to the inner workings of the evolutionary process, revealing how specific conditions interact with available genetic variation (whether from standing genetic variation or mutation) to influence the path taken. Data from other fields, especially molecular biology, can aid in the interpretation of the observed evolutionary paths, allowing us to develop mechanistic models to explain why certain outcomes should occur. Finally, the increasing abundance of biological data from all fields, and its organization into well-maintained databases, has made meta analysis an increasingly useful tool for developing, exploring and testing evolutionary hypotheses.

Throughout this thesis, I use experimental evolution with the model yeast *Saccharomyces cerevisiae* to study the processes of adaptation and speciation. My primary focus is on the underlying genetic changes, and that will be the focus of this chapter. Genetic systems are complex, and understanding how a genotype is translated into a phenotype is a central problem in biology. For this reason, the constraints and limitations that shape evolution are very difficult to predict or model *a priori*, and the genetic system is often the source of many of the surprises when conducting experiments.

The benefits and limitations of experimental evolution have been explored in previous reviews (e.g., BAILEY and BATAILLON 2016; BARRICK and LENSKI 2013; KAWECKI *et al.* 2012; LONG *et al.* 2015) and books (GARLAND and ROSE 2009), with some particularly focussed on microbes (ADAMS and ROSENZWEIG 2014; KASSEN 2014; LENSKI 2017). One major advantage of using a model organism such as *S. cerevisiae* for our experiments is that we have the ability to delve deeply into the underlying genetic basis of the observed outcomes, asking and answering genetic questions about evolution that would otherwise be very difficult to address. In this thesis, I hope to contribute to our collective understanding of evolution, especially as it relates to the genes involved in adaptation and how they might act given their specific genetic

and environmental context. I will explore how different factors can constrain and limit genetic adaptation and potentially contribute to speciation including: adaptive environment (both the qualitative nature of the environment and the exact concentration of a stressor), epistasis between adaptive mutations, ploidy level of an organism, and the interplay between these factors. These insights will help to inform future studies on organisms with more difficult to decipher genetic systems, including those from natural populations, by providing candidate explanations for observed phenomena that can be specifically targeted for testing.

In this Introduction chapter, I will first discuss genetic model organisms and their usefulness for the study of evolution as a backdrop for the rest of the thesis. I will then discuss the repeatability of adaptation and how genomic breadth (how many potential paths evolution might take) and mutation types (and their relative rates) might affect repeatability. Genetic interactions between potentially adaptive mutations can further shape the repeatability of adaptation if early adaptive mutations change the fitness effects of later mutations, directing evolution down certain paths that depend on those early stochastic events. These interactions result in lower repeatability than if all mutations had fixed effects, and also have implications for speciation among populations diverging in separate geographic areas (allopatry). Finally, I will discuss the limits of adaptation, as determined by the genomic breadth of adaptive mutations in a specific environment and by the availability of such mutations to the organism in question, before moving on to explore how large-scale datasets from molecular biology can potentially help us to interpret evolutionary data and hypothesize about possible evolutionary trajectories.

1.1 Genetic model organisms and the study of evolution

Model organisms are useful in the study of adaptation and speciation, especially when we are specifically interested in uncovering the genetic and molecular basis. By using the same model organisms for studies of evolution as for molecular biology, we can take advantage of the information and perspectives gathered from years of study by hundreds of scientists. For example, many genes have been characterized in these model organisms, and information about them has been organized into openly-available databases (yeast: *Saccharomyces* Genome Database, or SGD, CHERRY *et al.* 2011; *Caenorhabditis elegans*: WormBase, STEIN *et al.* 2001; *Drosophila melanogaster*: FlyBase, GRAMATES *et al.* 2017; *Escherichia coli*: EcoCyc, KESELER *et al.* 2017 and others; *Arabidopsis thaliana*: The Arabidopsis Information Resource, or TAIR, HUALA *et al.* 2001). In addition, these model organisms tend to have small, easily sequenced genomes with high-quality reference genomes, so candidate causative mutations can be found with relative ease. Because of the availability of molecular information, it is easier to filter this list of mutations by gene function to find those that are most likely to be responsible for the phenotype of interest. Finally, the many genetic tools and tricks that have been developed for these organisms simplify testing of specific genetic hypotheses.

While each organism can be used for a variety of questions, some are better suited for certain interests. For example, *Caenorhabditis* species are a good choice for studies about the evolution of sex, because one can manipulate the amount of outbreeding present (CUTTER 2005). *D. melanogaster* is well-suited to study the dynamics of evolution with obligate outcrossing (BURKE *et al.* 2010). And if an aspect of plant evolution is of interest, *A. thaliana* is the natural choice (MAURICIO 1998). Yeast is well-suited for studies where we

wish to characterize the nature of the genes involved in evolution, including but not limited to their identity. *S. cerevisiae* is the best genetically characterized eukaryote and has easily manipulated genetics (BOTSTEIN and FINK 2011). In yeast, it is standard practice to insert specific mutations into known locations of the genome by transformation and perform controlled crosses where individual meiotic progeny can be dissected out and grown separately (SHERMAN 2002). It was for these reasons, and because of their fast generation time for a eukaryote, that I chose to study yeast.

Yeast is an increasingly popular model system for experimental evolution and evolutionary genetics, including speciation research. Their rapid generation time (replicating in ~90 minutes under ideal conditions, SHERMAN 2002) allows for evolution experiments to be conducted efficiently. Yeast primarily reproduce asexually, either in a haploid or a diploid state, and can do so in either liquid or solid medium. While many strains of *S. cerevisiae* form clumps in liquid medium (e.g., brewing strains, SOARES 2011), lab strains have been selected to have dispersed cells (MORTIMER and JOHNSTON 1986), facilitating the quantification of abundance through the use of optical density measurements. Fast, asexual growth on agar allows for easy phenotyping and for bottlenecking of populations down to individual genotypes, the progeny of which can be used in fitness assays or for DNA extraction. Yeast can also be preserved in a frozen state, allowing for the long-term storage of strains, and their well-annotated, small genome (12 Mb) is easily sequenced (SHERMAN 2002). Although the 'natural' ecology of yeast is largely unknown, the many advantages of yeast as a genetic system have prompted studies into the evolution and ecology of natural yeast (as reviewed in LANDRY *et al.* 2006, REPLANSKY *et al.* 2008, LITI 2015). Studies of natural yeast complement lab investigations, developing yeast into a more complete model for studies of evolution.

1.2 Repeatability of evolution

Stephen Jay Gould proposed a thought experiment in his book *Wonderful Life* in which we might learn about the repeatability, and therefore predictability, of evolution by "replaying life's tape" and looking at the similarity of the outcomes (GOULD 1989). For natural populations, the closest we can come to replaying evolution is observing cases of parallel adaptation, where adaptation has "played" multiple times in separate but similar environments. From these populations, we know of many examples of phenotypic repeatability including the repeated loss of lateral plates when stickleback move from the marine environment into freshwater (SCHLUTER *et al.* 2004), the evolution of species of *Anolis* lizards with similar niches, morphology and behaviour (ecomorphs) on separate islands (LOSOS 1992) and the development of a similar gene expression pattern among clinical isolates of *Pseudomonas aeruginosa* from patients with cystic fibrosis (HUSE *et al.* 2010). Some cases of parallel phenotypic evolution are also underlain by parallel genotypic evolution, like wing pigmentation patterns in male *Drosophila* that involve regulatory changes of the same gene (*yellow*) (PRUD'HOMME *et al.* 2006) and stickleback morphological trait divergence in species pairs (CONTE *et al.* 2015), but others are not, like beach mice adapting to sandy coastal dunes in Florida's Gulf and Atlantic coasts (MANCEAU *et al.* 2010). I will focus the rest of this discussion on genetic repeatability, as that is the main focus of Chapter 2.

The generality of genetic repeatability in natural systems can be studied using meta-analyses, such as

that performed by CONTE *et al.* (2012). In this study, the authors estimated the probability of gene reuse for natural parallel phenotypic evolution and found that the mean probabilities were relatively high (0.32 for studies where the whole genome was considered when mapping phenotypes, 0.55 for studies where only candidate genes were tested for association with the phenotype). They concluded that the biases and constraints of evolution were quite strong, restricting the paths available to evolution and resulting in high repeated gene use. Gene reuse was also higher for more closely related populations, which can be interpreted as either a similarity in their genomic biases, in the standing genetic variation available to them, or a combination of the two. Truly parallel environmental conditions are difficult to verify in nature, however, where there are seemingly endless sources of ecological variation experienced by populations that could all affect the probability of certain evolutionary outcomes. If a pair of natural populations adapt differently to seemingly similar conditions, we cannot be sure that there was not a key unmeasured difference between their environments that caused their divergence, and this problem only increases with increasing numbers of populations (and, therefore, environments) included.

Experimental evolution allows us to "play life's tape" multiple times for a single set of conditions, concurrently, and compare the results. In such studies, different degrees of repeatability have been observed. In a study by MEYER *et al.* (2012), phage λ, which infect *E. coli*, were evolved and many developed the ability to utilize a new receptor, OmpF. All adaptive mutations were in the virus' J protein, and the phage that were able to use OmpF all had four mutations in common: two identical mutations across strains, a mutation in one of two adjacent positions and a mutation within a certain 30-basepair block. Another study by LOURENÇO *et al.* (2016) found seven genes targeted in parallel instances of adaptation of *E. coli* during gut colonization of mice. In yeast, a study by LANG *et al.* (2013) found a small subset of genes that drove increases in fitness in rich medium that were repeated among replicates (observed in 3 - 21 replicate populations out of 40). Similarly, in yeast adapting to a rich medium, KRYAZHIMSKIY *et al.* (2014) observed some genes to evolve repeatedly to varying degrees (each in 3 - 12 out of 104 total populations), and the proportion of gene use varied among starting genetic backgrounds. Many more examples of repeated evolution from experimental evolution studies have been reviewed in LOBKOVSKY and KOONIN (2012).

From these studies, we know that repeatability varies, but it is still unclear how it will vary depending on the environmental conditions. We know that it should vary with the genomic breadth of potentially beneficial mutations, which is itself tied to the environment in which adaptation occurs (e.g., GRESHAM *et al.* 2008). If many mutations are potentially beneficial in an environment, then there are many potential paths that evolution can follow. If, in contrast, only a few specific mutations will increase fitness of an organism in an environment, then adaptation will be highly repeatable due to this constraint. Repeatability should also vary with the types of mutations that are potentially beneficial and their relative mutation rates. For example, if both additional chromosomal aneuploidy and specific SNPs in a single gene increase fitness in an environment but aneuploidy occurs more frequently then we expect evolution to use aneuploidy more often because of the increased opportunity. In Chapter 2, I explore the effects of environment and mutation type on genetic repeatability in parallel populations adapted to high concentrations of copper. By testing individual factors (such as environment type) for their role in repeatability as well as considering genetic results within the context of broader biological knowledge (mutation types, molecular or biochemical characterization), we

can begin to understand the mechanics of repeated evolution.

1.3 Genetic interactions

Genetic interactions, measured as epistasis between alleles, also shape the outcomes, and therefore repeatability, of evolution. The word "epistasis" has a few different meanings in genetics and evolution, although these meanings are related (ROTH *et al.* 2009). Originally coined by BATESON (1909), epistasis was used to describe the interaction between genes in which the action of one gene was blocked by the action of another. This is how molecular geneticists still use the word, and they use epistasis analysis to determine the order of action of genes in regulatory pathways (AVERY and WASSERMAN 1992). Evolutionary geneticists, on the other hand, use the word to broadly describe any form of gene (or allele) interaction deviating from the expectation of independent effects (WHITLOCK *et al.* 1995). Depending on the appropriate scale for the phenotype being measured, either additive or multiplicative interactions can be the expectation.

In evolutionary biology, the ultimate phenotype of interest is usually fitness. Fitness, and traits often used as a proxy for fitness such as growth rate, are composite traits, underlain by many other biological processes. It is easy for epistasis to arise as a consequence of the many potential interactions between these underlying biological traits in determining the ultimate fitness of the organism. For example, if fitness depends on the products of a metabolic pathway and flux through that pathway is optimized at an intermediate value, the fitness effects of mutations in the enzymes of the pathway will depend on the flux rate of the whole pathway relative to the optimum (SZATHMÁRY 1993). If flux were below the optimum, a small effect mutation that increases flux would be beneficial. If there were a larger effect mutation that increases flux slightly above the optimum, this mutation may also be beneficial and spread in the population. But, the second mutation would make the first flux-increasing mutation deleterious in the new genetic background. This example is similar to any other trait undergoing stabilizing selection and will result in epistasis for fitness among loci affecting the trait, even if those loci act additively on the underlying trait (WHITLOCK *et al.* 1995).

The nature of epistasis between potentially beneficial alleles can determine the speed of adaptation or constrain evolutionary trajectories. Positive epistasis between beneficial alleles, where the two mutations together are fitter than the expectation, can theoretically increase the rate of adaptation, but few cases of positive epistasis have been found for beneficial alleles (BUSKIRK *et al.* 2017; CHOU *et al.* 2009). Negative epistasis is more commonly observed, where new beneficial alleles have smaller effects when in the same genetic background as other beneficial alleles, slowing the rate of adaptation (CHOU *et al.* 2011; KRYAZHIMSKIY *et al.* 2014). When epistasis becomes so negative that adding the second mutation is actually detrimental instead of beneficial ("sign epistasis"), the evolutionary trajectory of a population can be strongly constrained (WEINREICH *et al.* 2005). An example of this was found by TENAILLON *et al.* (2012) who evolved 115 populations of *E. coli* to high temperature. They found repeated evolution at the gene level, but there were at least two distinct pathways available, involving either the RNA polymerase complex or the termination factor rho. Early adaptive mutations were inferred to have directed later adaptation of the populations into one of these two paths through negative epistasis between these early mutations and subsequent ones. In this way, epistasis affects the amount of repeatability observed between populations since repeatability will be

lower when there are mutually-exclusive solutions for evolution to explore than when all beneficial mutations maintain their fitness effects regardless of genetic background. In the case of full independence, we would expect all populations to eventually converge upon the single best genotype, even with initial stochastic divergence.

1.3.1 Bateson-Dobzhansky-Muller model of speciation

The effects of epistasis not only shape adaptation within a population, but they also influence the outcomes of hybridization between populations in the main genetic model of speciation, the Bateson-Dobzhansky-Muller (BDM) model. In the BDM model of speciation, alleles that are beneficial or benign in their normal genetic background cause sterility or inviability in hybrid individuals (ORR 1995). Such interactions are known as BDM incompatibilities. The alleles are able to spread in different sub-populations because they do not cause reductions in fitness in those sub-populations' genetic backgrounds. Then, when the sub-populations are brought together and hybrids are formed, the alleles are tested together in novel genetic combinations and hybrid individuals have lower fitness than either of the parental types, reflecting negative sign epistasis. Low hybrid fitness results in decreased gene flow between sub-populations and, therefore, reproductive isolation.

Despite interest in BDM incompatibilities, only a few examples have been characterized at the molecular level (PRESGRAVES 2010). Most of these are found between species adapted to different local environments and the causative alleles are presumably beneficial in those separate environments (documented in NOSIL and SCHLUTER 2011). Other genetic incompatibilities depend on the environment in which they are measured, as have been found in natural populations of yeast (including one characterized two-locus BDM, HOU *et al.* 2015), but it is unknown whether the evolution of these alleles is linked to the external environment since we have no concrete knowledge of the evolutionary history. Additionally, some cases of incompatibility between natural populations have no clear connection to the external selective environment (see examples in MAHESHWARI and BARBASH 2011).

Experimental evolution studies allow direct control over the environment, and incompatible mutations have been found in some cases. DETTMAN *et al.* (2008) found reduced reproductive success in matings of lineages adapted to different environments compared to matings between lineages adapted to the same environment in *Neurospora crassa*, consistent with the action of BDM incompatibilities. They found similar results when mating populations of *S. cerevisiae* that had adapted to different environments (DETTMAN *et al.* 2007). In the latter case, the underlying BDM incompatibility was subsequently mapped (ANDERSON *et al.* 2010) and was the first reported BDM interaction among known genes that was isolated from experimentally evolved strains, to my knowledge. An incompatibility has also been found among experimentally evolved populations adapting to the same environment in *S. cerevisiae* (KVITEK and SHERLOCK 2011). There is another case of an experimentally evolved incompatibility arising from parallel selection in *Methylobacterium extorquens* (CHOU *et al.* 2014), but the nature of species (and reproductive isolation) is less clear for organisms without meiosis thus making it difficult to classify this interaction as a BDM incompatibility.

The relatively rapid (within 448 mitotic generations in the case of KVITEK and SHERLOCK 2011) establishment of alleles leading to BDM incompatibilities in the above experiments can be attributed to the accumulation of alleles driven to high frequency by selection. There are two main mechanisms of speciation

by selection, referred to as ecological and mutation-order speciation (SCHLUTER 2009). Ecological speciation occurs when the divergence between groups is driven by divergent selection in different environments (SCHLUTER 2009). Since selection is divergent, it drives the fixation of different alleles in each group, each advantageous in the local environment but not necessarily in the other. This process can lead to the evolution of any type of reproductive isolation, including premating isolation, hybrid sterility and intrinsic hybrid inviability as well as extrinsic, ecologically based pre- and postzygotic isolation (SCHLUTER 2009). In contrast, mutation-order speciation involves divergence between groups that occurs as a by-product of different mutations arising and fixing in separate groups adapting to similar selection pressures (SCHLUTER 2009). The evolution of reproductive isolation occurs by the chance fixation of different advantageous mutations in different groups, even though the same mutations would be initially favoured in both (SCHLUTER 2009). Early divergent adaptive mutations lead to increased divergence between groups due to the effects of epistasis influencing the evolutionary trajectories. It is important to emphasize that, even in identical environments, selection can become divergent between groups when the genetic background changes, thereby changing the mutations that confer a fitness benefit. Intrinsic reproductive isolation can then arise as a consequence of incompatible genetic solutions.

The aforementioned studies demonstrate that BDM-type interactions can establish over the course of hundreds of generations in experimental evolution studies by either mechanism of speciation by selection. What remains unknown thus far from long-term experiments of populations evolved under the same selective pressure is how frequently first-step adaptive mutations themselves could contribute to reproductive isolation. It is possible that incompatibility starts at the first steps of divergence in populations adapting in parallel, but it is also possible that those early stage mutations only change the selective environment of future mutations. I investigate epistasis between first-step adaptive mutations in Chapter 3 using fungicide-adapted mutations. I find that BDM incompatibilities are fairly common between these different, large effect adaptive mutations, demonstrating that mutation-order speciation may arise even at the very first step of adaptation.

1.4 Limits to adaptation

In extreme cases of evolutionary constraint, populations will not be able to respond genetically to the environment and will have reached their adaptive limit. Such adaptive limits are thought to contribute to the limited geographic distribution of species, and have been suggested by observed trade-offs between traits important for adaptation to different edges of a species' range (for example, ANGERT *et al.* 2008). The limits to adaptation include not only whether a beneficial mutation is possible, but also whether it will be able to successfully rise in frequency. This depends on a host of population genetic factors, most of which I will not address here, but have been well reviewed in BARTON and PARTRIDGE (2000). In this section, I will primarily focus on the availability of genetic variation for selection to act upon. Lack of genetic variation in the trait of interest is a common explanation for the observation of a limit to evolutionary change (BELL 2013; BLOWS and HOFFMANN 2005). Most papers on the topic are primarily concerned with natural populations, in which standing genetic variation is the primary source of variability. In experimental evolution studies, unless standing genetic variation is of explicit interest, the populations are typically started with as

little variation as possible, coming from either inbred lines or originating from single clones. In these experiments, genetic variation still determines the rate of initial response to selection (BARTON and PARTRIDGE 2000), but it is limited by the input of spontaneous mutation (DE VISSER and ROZEN 2005).

When considering a specific trait under selection, or a specific environment to which a population is trying to adapt, the beneficial input of mutation can be limited in a few different ways. One well-discussed limit is imposed by mechanistic, physiological or developmental constraints of the organism, where certain types of variation are not possible (BLOWS and HOFFMANN 2005; SMITH *et al.* 1985). The same epistatic relationships that constrain evolution to follow certain trajectories dependent upon initial beneficial mutations can also limit further evolution entirely in extreme cases (SMITH *et al.* 1985). This arises as a result of sign epistasis between the current genetic background and all potential beneficial mutations, making these mutations inaccessible. These factors affect the mutational target size for adaptation. In a given environment, if the genomic breadth of potentially beneficial mutations is non-existent, the organism in question will not be able to adapt and will have met its adaptive limits. Another possibility is that a population is instead limited by the mutation rate (BLOWS and HOFFMANN 2005). If potentially beneficial mutations are possible but are exceedingly rare due to very low mutation rates for those specific genomic changes, it is very unlikely that a population will be able to adapt. It will have effectively met its adaptive limits and, if it has a negative growth rate, will go extinct before being rescued by evolution. Because the genomic breadth of adaptation and the rate of beneficial mutation are decided by the environment in question and the biology of the organism, changing either can cause an organism to reach its adaptive limit. In Chapter 4, I look at ploidy as an intrinsic biological feature that can affect the accessibility of beneficial mutations. Ploidy is not often considered among the factors thought to influence the limits of adaptation, but both the effects of mutations (related to genomic breadth) and the rates of certain mutations (especially certain mutation types) can differ by ploidy, even in otherwise identical genomic backgrounds. It is thus important to determine how ploidy can affect evolutionary dynamics, especially when considering the evolution of antibiotic resistance, for which both haploids and diploids are common targets.

1.5 What can large-scale datasets tell us about evolution?

Forward genetic screens have been used to determine the genetic basis of phenotypes since the 1940s (BEADLE and TATUM 1941). In these studies, collections of mutants, usually generated by a mutagen, are screened for a phenotype of interest thought to be related to a particular biological process. All mutants displaying the phenotype of interest are characterized, both genetically and phenotypically. Genetic characterization can involve sorting the mutants into complementation groups (where each group is likely to represent a single gene), genetic mapping of causative mutations, and sequencing of those mutations. Phenotypic characterization depends on the biological process of interest, but could involve morphological measurements, determination of cellular localization patterns, and biochemical measurements of underlying reactions. Much of our knowledge about the functions of individual genes was collected in this way. Then, to order these genes into genetic pathways leading to the phenotype of interest, double mutants were created and epistatic analysis was performed. In molecular genetics, this type of work has been complemented by the use of

mutation collections and large-scale screens, which has greatly expedited the collection of biological data. In this section, I would like to explore what we, as evolutionary biologists, might learn from these large-scale datasets. This data has been primarily applied to models predicting the evolution of microbial systems and has been reviewed elsewhere (PAPP *et al.* 2011), but I would like to broaden the view to include other natural systems. Specifically, I am interested in the genetic interaction datasets, which tell us about epistasis between large collections of mutations. I see there being three major barriers to using this kind of data to understand evolution. First, these experiments are often performed with knockout mutations, but does this kind of mutation reflect common natural mutations? Second, how generalizable are the specific interaction results across different levels of biology (different populations, genetic backgrounds, species, genera, etc.)? And third, how do these genetic interactions depend on the environment in which they are measured? In the following sections, I consider different data sets that provide insight into the answers to these questions.

1.5.1 How well do mutation collections represent natural mutations?

Each of the model genetic organisms explored above has a large-scale mutation collection (yeast: deletion collection, GIAEVER *et al.* 2002; *C. elegans*: deletion collection, C. ELEGANS DELETION MUTANT CONSORTIUM AND OTHERS 2012, and million mutation project, THOMPSON *et al.* 2013; *D. melanogaster*: gene disruption project, BELLEN *et al.* 2011; *E. coli*: Keio deletion collection, BABA *et al.* 2006; *A. thaliana*: T-DNA insertion collection, ALONSO *et al.* 2003). The collections differ in scope (proportion of the genes covered) and mutation type, but perhaps the best collection is in yeast. The large-scale whole-gene deletion collection in *S. cerevisiae* includes all viable gene deletions and all essential genes deleted in a heterozygous state (GIAEVER *et al.* 2002), as well as temperature-sensitive alleles for the essential genes (COSTANZO *et al.* 2016). In addition, there are collections of plasmids available in yeast for overexpression studies (high copy plasmids) (JONES *et al.* 2008) and complementation mapping of genes (low copy plasmids) (HO *et al.* 2009; HVORECNY and PRELICH 2010). Together, these collections have been used to characterize general properties of the cell (e.g., finding the number of essential genes) as well as the functions of individual proteins and their interactions, but how well do these mutations represent naturally occurring mutations?

A study by DOWELL *et al.* (2010) found that, even for whole-gene deletions, the effect of some mutations can depend on their genetic background. When they systematically deleted genes in a second strain of *S. cerevisiae*, 57 genes were only essential in one of the two backgrounds (DOWELL *et al.* 2010), indicating that results from large-scale deletion collections may not be universally applicable, especially for less severe phenotypes. However, a study by PAYEN *et al.* (2016) used single-gene deletions and amplifications to identify potentially beneficial mutations in three environments (limitation of phosphate, glucose or sulfate) by their fitness effects in pooled competitions. When comparing the results to sets of mutations found during experimental evolution in the same conditions, they found that, on average, ~35% of the mutations that occurred during experimental evolution were predicted to be beneficial by the systematic screen (PAYEN *et al.* 2016). In addition, a study by JELIER *et al.* (2011) found that they were generally able to predict the phenotypes of a variety of yeast strains based on their genome sequences. They first predicted the impact of individual mutations within genes on protein function, focussing on loss-of-function mutations. Then, utilizing high-throughput data sets containing the growth rates of the deletion collection under different

environmental conditions acquired from the SGD (CHERRY *et al.* 2011), they estimated the relevance of each protein perturbation to growth in each condition. From this, they predicted the growth of each strain in each condition based on the strain's mutated genes, and they assessed the performance of their predictions by measuring growth across 20 conditions.

The ability to predict the potential fitness effects of loss-of-function type mutations *en masse* is amazing in and of itself, but can we do the same kind of analysis without limiting it to predicted loss-of-function mutations? I have attempted to begin to investigate this question by taking sets of genes found to impact resistance to chemicals from EHRENREICH *et al.* (2012) and determining whether these genes also impact fitness, on average, when deleted as either a heterozygote or homozygote, in a study by LEE *et al.* (2014). EHRENREICH *et al.* (2012) used a technique that they termed "extreme QTL mapping" to identify the genetic basis of resistance to 13 chemicals in segregants of all pairwise crosses of four ecologically and genetically diverse yeast strains. In doing so, they detected more than 800 loci with an effect in at least one of the drugs. From their set of 13 chemicals, I identified five chemicals that were also tested in the screens of LEE *et al.* (2014), one of which had been tested twice at two different concentrations. LEE *et al.* (2014) tested ~1,100 yeast strains heterozygous for a deletion in an essential gene and ~4,800 yeast strains homozygous for a deletion in a nonessential gene for their response to 3,250 chemicals that inhibit wild-type growth.

If both natural mutations and gene deletions uncover roles for genes in certain environmental conditions, then we would expect the same genes to be implicated in fitness deviations in both studies. I will use the candidate genes from EHRENREICH *et al.* (2012) to determine which genes out of the deletion set of LEE *et al.* (2014) are predicted to have large fitness deviations in certain chemicals. For all chemicals that were tested in both experiments, I retrieved the list of candidate genes that putatively affect resistance to that chemical from EHRENREICH *et al.* (2012). I then downloaded all of the fitness data from LEE *et al.* (2014) for those chemicals. Because fitness data was reported as either negative or positive (having a growth defect or growth advantage in the chemical when compared to the control), I took the absolute value of all fitness deviations to test only whether the genes had an effect, ignoring the direction of that effect. This should include cases where the deletion of a gene has a negative effect on fitness but an overexpression allele of the same gene (which is possibly present among the natural variants) has a positive effect. In order to determine whether candidate genes had, on average, a larger effect on fitness than expected in a certain chemical, I took 10,000 random samples (without replacement) of the same size as the number of candidate genes from the full dataset for that chemical. I then compared the mean fitness deviation of the candidate genes to the distribution of means calculated from the samples. I found that the candidate genes only had a larger average effect than expected by chance (>95% of samples) in two cases: in 4-nitroquinoline 1-oxide (4-NQO) and in one of the two experiments run with tunicamycin (Fig. 1.1). Of the two concentrations of tunicamycin tested in LEE *et al.* (2014) (25 nM and 200.47 nM), the candidate genes had a larger fitness effect compared to the average in the higher one (200.47 nM), which is closer to the concentrations used by EHRENREICH *et al.* (2012) for their selective plates (between 2.5 μM and 3.5 μM).

From this analysis, we conclude that genes responsible for natural variation in an environment do not correspond well with those whose deletions show the largest fitness response in that environment. This lack of correspondence can be partially attributed to the differences in methods between the two studies

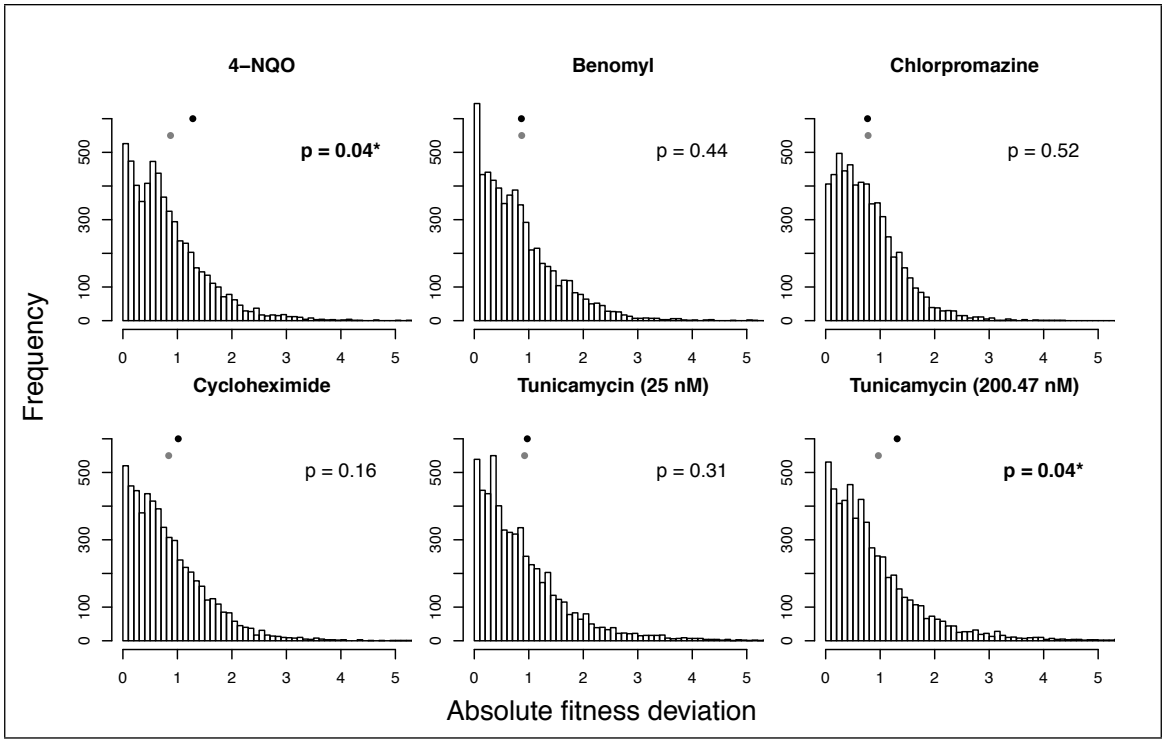

Figure 1.1: Histograms of the distributions of absolute fitness deviations in each chemical investigated (data from LEE *et al.* 2014). X-axes are cut off to exclude extreme values but include at least 99% of the data. Above each histogram, the means of the whole dataset (in grey) and the candidate genes for that chemical from EHRENREICH *et al.* (2012) (in black) are shown. P-values are calculated as the proportion of randomly sampled means (out of 10,000 samples) that are greater than the mean of the candidate genes in that chemical (one-tailed test). Significant p-values are in bold with an asterisk (*).

being compared. Fitness was assayed differently in the two experiments, where EHRENREICH *et al.* (2012) used plates containing chemicals to select for highly resistant segregants and LEE *et al.* (2014) extracted fitness data from yeast growing in liquid medium. LEE *et al.* (2014) grew pooled samples of yeast in liquid medium that contained the chemical of interest, where each pool consisted of either all heterozygous or all homozygous strains. Because each deletion is uniquely barcoded, the fitness defect of each strain in the pool was measured by the relative abundance of its barcode in sequenced pools from the treatment samples compared to control samples (without chemicals), allowing for a quantitative measurement of fitness. The different types of medium used between studies, in addition to differing concentrations of chemical used both within the study done by EHRENREICH *et al.* (2012) (different selective concentrations were used for different crosses) and between the two studies, could potentially select for mutations in different genes. This will be a persistent problem in using molecular biological data to interpret evolution, however, because we cannot expect methods to be the same unless the studies are conducted for the purpose of comparison.

In addition, due to the nature of QTL mapping, EHRENREICH *et al.* (2012) could not identify the causative gene in most cases, but instead mapped the phenotypes to small windows containing up to 14 candidate genes, after excluding genes that contained no segregating polymorphisms. This means that the genes included in our analysis may not be causative in the original crosses of EHRENREICH *et al.* (2012). Again, this will be a persistent problem with QTL analyses, where we might have to restrict ourselves to using large-scale data to help narrow down candidate gene sets to those most likely to be involved. The candidate genes from EHRENREICH *et al.* (2012) were also mapped based on naturally-occurring variation, as opposed to the deletion mutations used in LEE *et al.* (2014). This was the key comparison that I was interested in making because evolutionary data will consist of these naturally-occurring mutations, but we expect some types of natural mutations (like gain-of-function alleles) to be especially poorly represented by deletion mutations. The mutations from EHRENREICH *et al.* (2012) were also mapped in haploid populations, as opposed to the diploids used in LEE *et al.* (2014). If mutations only have an affect on the phenotype in one of the two ploidies, they will only be observed in one of the two screens. Additionally, for all essential genes (those that cannot be fully knocked out), only heterozygous mutants were tested in LEE *et al.* (2014), while haploid mutations in these genes could exist among the set screened by EHRENREICH *et al.* (2012), if non-knockout mutations are tolerated. I also included candidate genes found in any genetic background by EHRENREICH *et al.* (2012), not limiting the set to those found in the BY background, which is shared by the deletion mutants of LEE *et al.* (2014). Based on the poor predictive power observed in our comparison within a single species, it seems unlikely that we can currently use single large-scale datasets to predict phenotype-causing alleles in other, more phylogenetically distant, organisms.

Regardless, it is encouraging that I found any correspondence at all in this small comparison, and the results of JELIER *et al.* (2011) give me confidence that collecting more of these large-scale datasets and performing more of these kinds of comparisons will lead to a better ability to infer phenotypic causality from genomic data of phenotyped individuals in the future. Although, in this analysis, I have used mapped, natural variants to predict the phenotypes of deletion mutants, I imagine that this kind of comparison could be done in the opposite direction to identify candidate natural variants involved in producing a phenotype of interest. If, for example, we were trying to determine the alleles involved in copper tolerance of an organism

and we had the relevant genomic data, we might look at a dataset of deletion mutant phenotypes in copper, like those produced by LEE *et al.* (2014), and identify the deletions with the highest fitness deviations. If any of the corresponding genes are mutated in our organism's genomic data (especially when compared with a close relative that differs in its copper tolerance), then those genes might be implicated in copper tolerance for this organism. This method could be used as a supplement to QTL mapping or to inform candidate gene analysis. We only based our results on one set of data, unlike JELIER *et al.* (2011) who had multiple datasets available to them per environment in some cases. For these environments, they were able to choose the more reliable dataset, based on the connectivity of the implicated genes in a predicted functional gene network for yeast (YeastNet, LEE *et al.* 2007). By continuing to collect high-throughput datasets, we can improve on the quality of information available for the phenotypic effects of genic mutations. It would be especially useful if future high-throughput studies collected data from other types of mutations (such as using overexpression plasmids) to get a sense for how different mutations within a single gene can differ in their effects. In addition, a deeper understanding of the nature of the proteins underlying a trait, and the domains of which those proteins are composed, may help inferential power in the future. Mutations that are not likely to cause loss-of-function of the whole protein may instead only affect a single protein domain and may have specific, as opposed to general, effects on protein function, interactions and phenotype (RYAN *et al.* 2013). While current tools do not have the ability to predict the effects of such mutations for proteins as a whole (RYAN *et al.* 2013), future advances in functional protein prediction used in conjunction with information from a variety of large-scale assays may enable better mapping of mutations to phenotype.

1.5.2 How consistent are genetic interactions across levels of biological diversity?

In evolution, genetic interactions (or epistasis) determine the outcome and pace of adaptation, as well as the potential for genetic reproductive incompatibilities and speciation. From experimental work, we know that beneficial mutations often behave non-additively, at least for fitness (e.g., my Chapter 3, BUSKIRK *et al.* 2017; CHOU *et al.* 2009, 2011; KRYAZHIMSKIY *et al.* 2014). Molecular biologists also have a long history of studying genetic interactions, traditionally by using suppressor screens. Now, large-scale genetic interaction studies are being performed that measure epistasis on a genome-wide scale. In organisms where it is feasible, like *S. cerevisiae* (COSTANZO *et al.* 2016) and *Schizosaccharomyces pombe* (ROGUEV *et al.* 2008), interactions are measured by assessing the growth of double mutant strains. These double mutants are generated in high-throughput experiments using genetic systems developed specifically for this purpose, allowing for the simultaneous generation and measurement of thousands of strains. In other organisms, like *C. elegans* (TISCHLER *et al.* 2008), RNA interference (RNAi) is used to knock down expression of a gene, simulating the effect of a null mutation. RNAi can either be used combinatorially (where two genes are simultaneously targeted), or in combination with a homozygous mutant animal. I will refer to these as double mutants throughout to simplify explanation. Once double mutant organisms have been created, genetic interactions are detected by deviations of the double mutant phenotype from that which is expected based on the effects of the two mutations in isolation. The phenotype being measured is usually a proxy for viability or growth and negative interactions are referred to as either synthetic sick (where the double mutant is less fit than expected, but still viable) or synthetic lethal (where the double mutant is inviable

despite both single mutants being viable). These studies have elucidated some general patterns about gene interaction networks, such as the existence of 'hub' genes, which interact with many more genes than average in the network, and that functionally related genes can be predicted based on the similarity of their interaction profiles (DIXON *et al.* 2009). Can evolutionary biologists take advantage of this wealth of genetic interaction data to learn about the nature of epistasis in shaping evolutionary trajectories and make predictions about possibly interacting alleles in wild populations or in other species?

Large-scale genetic interaction screens have been performed in a few model organisms, but the largest and best-described dataset is for budding yeast, *S. cerevisiae* (COSTANZO *et al.* 2016). When large-scale interaction studies have been performed in other eukaryotic systems, such as the fission yeast *S. pombe* and nematode worm *C. elegans*, the results are often compared with those from *S. cerevisiae*. In those comparisons, the global genetic network properties such as degree of interconnectedness and amount of crosstalk between specific biological processes tend to be well-conserved but individual genetic interactions are less conserved (DIXON *et al.* 2009; RYAN *et al.* 2012). In *S. pombe*, ~17-30% of negative interactions are conserved with *S. cerevisiae* in any individual experimental study (when ignoring gene function, see below) (DIXON *et al.* 2008; ROGUEV *et al.* 2008). DIXON *et al.* (2008), along with performing a large-scale interaction experiment, also curated the literature at the time for reported *S. pombe* genetic interactions (note that this set excludes the two previously mentioned experimental studies) and found that 18-23% of those interactions were conserved with *S. cerevisiae* (DIXON *et al.* 2008). In *C. elegans*, less than 5% of interactions are found to be conserved with *S. cerevisiae* (BYRNE *et al.* 2007; TISCHLER *et al.* 2008).

To see how well-conserved interactions might be over multiple species comparisons, I have downloaded the most recent version of the Biological General Repository for Interaction Datasets (BioGRID) genetic interaction dataset (release 3.4.151, CHATR-ARYAMONTRI *et al.* 2017) and chosen the species with the most non-redundant genetic interactions (>1500; *C. elegans*, *D. melanogaster*, *E. coli*, *Homo sapiens*, *S. cerevisiae*, and *S. pombe*) curated from large-scale screens as well as small-scale experiments. For these organisms, I found all available pairwise orthologs using Ensembl's BioMart tool (Ensembl release 90, AKEN *et al.* 2017, BioMart: KINSELLA *et al.* 2011), which didn't include *E. coli* or *S. pombe*. I excluded *E. coli* from analysis but, for *S. pombe*, I downloaded the set of manually curated orthologs between *S. pombe* and *S. cerevisiae* from PomBase (MCDOWALL *et al.* 2014). For all other species except *S. cerevisiae*, the fission yeast orthologs were determined by taking the list of budding yeast orthologs with that species and finding the fission yeast ortholog for the budding yeast gene. Note that I did not exclude orthologs that corresponded to sets of paralogs in the other species, so the total number of orthologous genes considered was not equal between both species of a pair. Only genetic interactions were considered for this analysis, but both positive and negative interactions were included. For each species pair, I first determined all (non-redundant) pairwise genetic interactions in each species between genes that had orthologs in the other species. I then compared these lists of interactions between the two species of a pair, checking for overlap where both interacting genes were orthologous to interacting genes in the other species. The results are plotted in Fig. 1.2.

Assuming that interactions are tested randomly with respect to gene identity and to the existence of an ortholog in another species, we can determine the proportion of interactions that we expect to be conserved

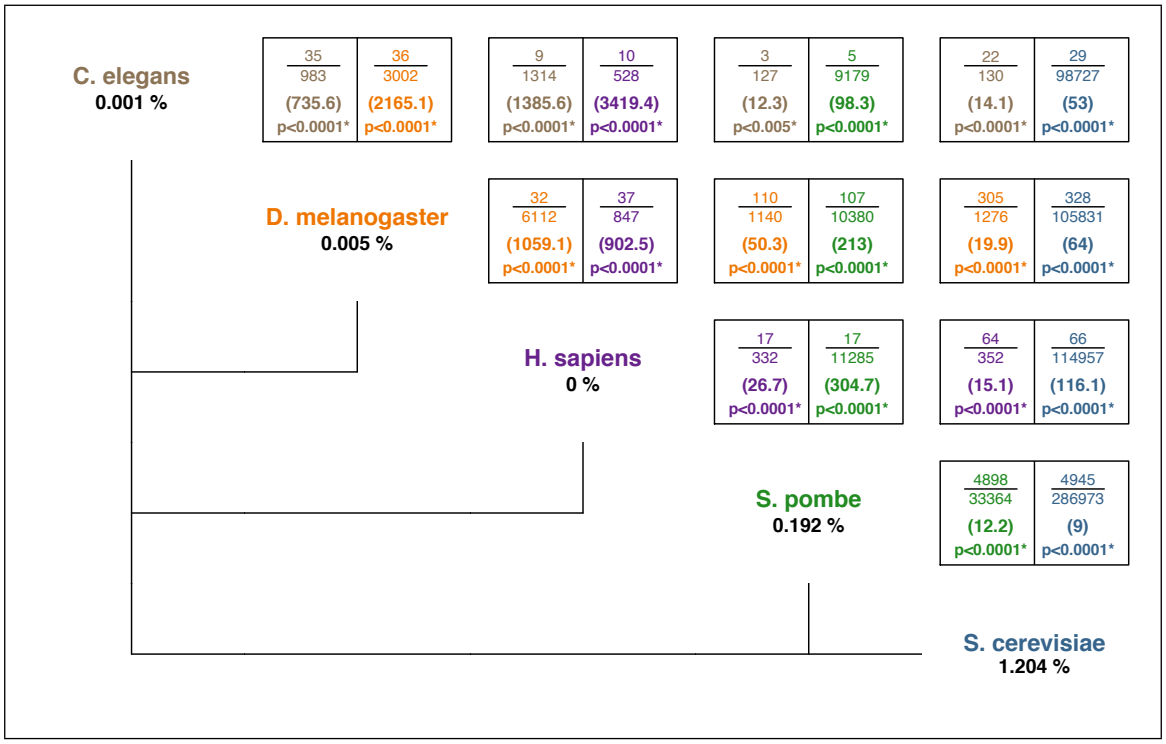

Figure 1.2: Conserved genetic interactions between pairs of species. Each pairwise combination of species is represented by a single box, found in the corresponding row and column to the species names. Below the species name is the percentage of total possible pairwise gene combinations that have been shown to have a significant interaction. The data for each species is coloured by the colour of the species name. Within each box, for each species, the number of conserved genetic interactions is above the line with the number of genetic interactions between genes of that species that have orthologs with the other species below the line. This is followed by the enrichment of conserved interactions observed compared to expected in parentheses and the p-value of a Binomial test where the null probability is the proportion of total possible gene pairs that are known to interact in the comparison species. Note that because we included paralogous genes, the number of conserved interactions does not necessarily match between the two species in the pair. Below the diagonal is a representation of the phylogenetic relationships between species (not to scale, created using phyloT: `http://phylot.biobyte.de/index.html`).

between species. For species A, we expect the conserved proportion of its interactions between its genes that have orthologs in species B to be the proportion of all possible pairs of genes for which a significant interaction has been found in species B (number of gene pairs tested and determined to interact significantly/total number of gene pairs = p). Data on the number of non-redundant gene pairs found to have a significant interaction comes from BioGRID (CHATR-ARYAMONTRI *et al.* 2017). The total number of gene pairs was calculated as the square of the total number of protein-coding genes. We excluded dubious genes from these numbers, which gave: *C. elegans*: 20,222 (release WS260, STEIN *et al.* 2001), *D. melanogaster*: 13,931 (release R6.17, GRAMATES *et al.* 2017), *H. sapiens*: 19,836 (GENCODE release v27, HARROW *et al.* 2012), *S. pombe*: 5,064 (release version 30th Jan 2017, MCDOWALL *et al.* 2014), and *S. cerevisiae*: 5,892 (genome inventory as of 1/18/2017, CHERRY *et al.* 2011). I used the expected proportions to perform Binomial tests to determine whether there were more conserved interactions than expected for each species in each species pair. For example, in the *C. elegans* - *D. melanogaster* comparison, I first found all significant interactions between genes in *C. elegans* that have orthologs in *D. melanogaster* (a total of n interactions). I then determined how many of these interactions were conserved (X conserved interactions where orthologs of both genes interact in *D. melanogaster*). The Binomial test was then performed using the total number of interactions (n) as the number of trials, the number of conserved interactions (X) as the number of successes and the expected proportion of conserved interactions in *D. melanogaster* (p) as the probability of success. I found that there were significantly more interactions conserved than expected in all cases (Fig. 1.2), even among the distantly related species.

I find that, when searching for orthologous genetic interactions, there is low conservation between species in general. Even in comparisons with *S. cerevisiae*, where most interactions have been studied, only 14.7 - 23.9% of interactions are conserved. When I attempt to account for sampling effort by dividing the fraction of conserved interactions by the expected proportion based on the total number of significant interactions known in the other species to get the enrichment of conserved interactions, however, I find that the conservation is much higher than what is expected by chance (Fig. 1.2, numbers in parentheses; all Binomial tests are significant). Conservation generally shows a pattern of decreasing enrichment with increasing phylogenetic distance and is very high between the animals (*C. elegans*, *D. melanogaster*, and *H. sapiens*). The enrichment between the yeasts, however, is much lower than the enrichment found between the animals or between each yeast species and each of the animals. It is possible that interactions found within multicellular animals are more likely to be conserved with other multicellular animals because of conserved specialization of orthologous genes' functions in developmentally complex organisms. In the yeasts, however, the genes may be less specialized and have more redundancy in function, leading to fewer specific (and therefore conserved) genetic interactions.

There are some important caveats to this analysis, however, that may lead to systematic overestimation of enrichment and underestimation of p-values. If there is a bias towards testing interactions between conserved genes (i.e., those most likely to have orthologs in other species) then we may be biased towards finding conserved interactions. We know that interactions are not tested randomly, with some studies focussing on particular biological processes (e.g., BYRNE *et al.* 2007 targeted genes in signal transduction pathways), and it is common practice to test a smaller number of 'query' mutations in combination with a

larger array of mutations, thus biasing the information gathered in favour of the chosen query genes. Such a bias could explain the difference in enrichment observed between the animals and the yeasts. Because fewer interactions have been tested in the animals, they are more biased towards interactions that are more likely to be conserved. There is a much greater number of interactions that are known in the yeast species, however, with almost all possible interactions having been tested in *S. cerevisiae* (COSTANZO *et al.* 2016), although many essential or nearly essential genes remain to be tested. Thus, there is less bias and a number of conserved interactions that is closer to the expectation. Additionally, we allowed each gene in our analysis to have multiple orthologs in the other species, but we have not accounted for these additional chances of finding a match in our Binomial test. By restricting our analyses to gene sets that have orthologs in all species being compared and that have been tested for interaction in those species, in addition to either using only orthologs with a single match in the other species or using probabilities weighted by the homology relationships, we could improve our ability to compare conservation between species pairs. Further, more studies covering larger, and less-biased, portions of the genome will allow for better estimates of the true enrichment of conserved interactions. An additional complication in using sequence-based homology relationships to assess conservation is that they do not always correspond to functional homology. For example, KACHROO *et al.* (2017) found that two genes in yeast can be replaced by non-orthologous genes from *E. coli*. If anything, these types of mismatches between sequence and functional orthology should make estimates of conservation more conservative when based on sequence data alone.

In addition to caveats to the analysis, there are other potential problems in using this kind of data for evolutionary interpretations. Most of the interaction data, and especially data from large-scale interaction studies, comes from loss-of-function mutations (either gene deletions or RNAi knockdown), which may not represent all natural mutations (see Section 1.5.1). In addition, the resulting alleles are generally either mildly deleterious or neutral. Interactions between deleterious mutations may differ in a biologically significant way from interactions between beneficial mutations. Also, because most large-scale genetic interaction mapping studies have used proxies for growth rate as the phenotype of interest (DIXON *et al.* 2009), there is not a lot of information about how interactions may differ in kind depending on the phenotype being measured. However, the focus on growth rate enables us to interpret the results in terms of fitness. Because of these issues and those of incomplete and biased data, the presence of an interaction in curated data may be used as evidence that epistasis is occurring between mutations in two different genes but the absence of interaction data should not be similarly taken as evidence of no epistasis. These datasets could be especially useful for mapping BDM incompatibilities between species where candidate genes could be found that are known to have synthetic sick or synthetic lethal interactions in a closely related model organism.

Despite low conservation of individual connections, larger-scale conservation can be useful in predicting functional evolution. Between *S. cerevisiae* and *S. pombe*, RYAN *et al.* (2012) found that, within protein complexes, interactions were well conserved (positive interactions: 70%, negative interactions: 68%). In this same comparison, as the level of biological association between the genes decreased, they found decreasing conservation of specific interactions (same biological process, positive: 58%, negative: 38%; separate processes, positive: 19%, negative: 15%) (RYAN *et al.* 2012). Also in comparisons between *S. pombe* and *S. cerevisiae*, similar percentages of conserved interactions were found by ROGUEV *et al.* (2008), and the

general patterns were supported by FROST *et al.* (2012). The frequency of interaction between biological processes is highly conserved, however, even if the exact connections are not (RYAN *et al.* 2012). In addition, highly connected hub genes found in *C. elegans* are also highly connected in other animals, suggesting conservation of certain genes as hubs across species (LEHNER *et al.* 2006). In light of these results, if we knew the biological processes underlying adaptation to a given environment, we may be able to predict the amount of epistasis that would be found between potentially adaptive alleles in that environment by using data on genetic interactions within and between those processes. This information could indicate how quickly evolution might be expected to proceed as well as whether it might be constrained to certain genetic trajectories. Similarly, we might be able to hypothesize about the amount of negative epistasis expected to be present between populations adapted to different environments based on the processes important for those environments. Further developments and increased data in this field will help to inform hypotheses about specific interacting alleles and determine the general patterns that we can expect to hold true for all organisms, elucidating how this knowledge can be applied to natural systems.

1.5.3 How do genetic interactions depend on the environment?

If we want to use large-scale genetic interaction data to help us understand evolution, an important consideration is how the observations are dependent upon the environment. Experimental studies are generally performed under a single set of standard laboratory conditions, thus not providing information about environment-dependence. In Chapter 3, I give an example of epistatic relationships changing with changing concentrations of a single stressor. These kinds of changes may be somewhat predictable, but probably only in the special case of changing severity of the evolutionary environment in which the mutations were selected (see **Discussion**). Another study on reproductive isolation in yeast (HOU *et al.* 2015) tested many intraspecific crosses on different culture conditions. All offspring were chosen to have high viability in the rich lab medium (YPD), but the authors found that the hybrid viability was environment-specific and varied among crosses. These results indicated that the underlying genetic incompatibilities were dependent upon the test conditions. Other evolution-based studies provide case studies showing that epistasis can depend on the environment in which it is measured (e.g., REMOLD and LENSKI 2004; WANG *et al.* 2009), but few directed investigations of the large-scale generality of the phenomenon have been performed.

A few studies have attempted to address this issue, primarily in *S. cerevisiae*. JASNOS *et al.* (2008) measured the maximum growth rate of single and double gene deletion strains in a benign environment and several stressful environments, finding that epistasis became positive (alleviating), on average, in more stressful environments. They attributed this change to the general properties of decreasing growth rate in stressful environments; when growth rate is impaired, other defects caused by mutations have less of an impact. They did not focus on how individual interactions changed with the environment, however, or whether individual interactions changed in a predictable way. While alleviation of deleterious effects may be the overall pattern, it is possible that individual interactions will differ, potentially in an important way. In addition, JASNOS *et al.* (2008) specifically chose to investigate deleterious mutations, but it is possible that neutral or beneficial mutations could show altogether different patterns. In addition, both targeted (ST ONGE *et al.* 2007) and broad (BANDYOPADHYAY *et al.* 2010) surveys of genetic interactomes in the presence and

absence of DNA-damaging agent methyl methanesulfonate (MMS) in yeast indicate that interactions can be environment-dependent. These studies indicate that changes in interactions may be somewhat predictable because the genes that had many changes in their interactions were often ones that were known to be sensitive to MMS when knocked out or genes with a role in DNA repair (BANDYOPADHYAY *et al.* 2010), and similar results were found in *E. coli* (KUMAR *et al.* 2016). Again in yeast, deletion mutations of paralogs from a whole genome duplication were also found to have altered patterns of epistasis under different experimental conditions (MUSSO *et al.* 2008). Paralog double mutants that were sensitive to certain stressful conditions were generally found to have functions related to that condition (MUSSO *et al.* 2008). Finally, condition-dependent epistasis is also predicted from models of metabolic networks in yeast (BARKER *et al.* 2015; HARRISON *et al.* 2007) and other microbes (JOSHI and PRASAD 2014), and a small number of the predicted interactions in yeast from HARRISON *et al.* (2007) were verified *in vivo*.

Further information is needed in this field, especially from experimental data. It would be relatively simple and worthwhile to test double mutant strains made in rich medium conditions in a variety of other conditions, as long as the mutants are viable in the original conditions. One generally ecologically-relevant trait that would be useful to test is temperature-dependence. There is a lot of interest in the effects of temperature right now as it relates to climate change and varies both geographically and temporally. For lab-reared organisms, testing for the effect of temperature on a mutant phenotype could be as simple as rearing the organism in a temperature-controlled incubator. I would generally expect growth to worsen in temperatures that are increasingly divergent from the optimum (due to the instability of the underlying proteins or slowing of enzymatic reactions), but especially large changes in phenotype could be examined to determine the underlying causes. From these kinds of experiments, I would hope to find out whether network properties remain stable, in general, and whether changes in environment have predictable and/or consistent effects on epistasis. These results would again have implications for how adaptation might proceed after an environmental change.

Changing epistasis in changing environmental conditions also has implications for how we conceptualize reproductive isolation in speciation. Reproductive isolation builds up from a combination of isolating barriers, which have been traditionally categorized as either extrinsic or intrinsic. Extrinsic barriers arise as a mismatch of the organism's phenotype to the environment in which it is found. Common examples include immigrant inviability and hybrid partial inviability due to intermediacy of phenotype in one of the selective environments. Genetic incompatibilities, on the other hand, have been traditionally classified as intrinsic barriers, independent of the environment. If the incompatibility of alleles often depends on the environment in which it is measured, then these 'intrinsic' isolating factors will have an extrinsic basis and we will have to include these kinds of cases in our models of speciation. If hybrids between species are likely to find themselves in qualitatively different environments from their parents, knowing how epistasis changes with environmental conditions will be especially important for the stability of a reproductive barrier.

1.6 Summary

Despite the many recent advances in our knowledge about the genetic basis of evolution, there is still much to be discovered. Experimental evolution using model organisms is helping to lead the way in the molecular dissection of evolution, informing us about the repeatability of evolution under different circumstances, and how that repeatability depends on the genetic properties of the organism and their capacity for adaptation. In addition, we have gathered evidence on the profound effects that epistasis can have on both adaptation and speciation. These insights have allowed us to begin to interpret the abundant genomic data accumulating in natural systems, but I believe that there is still more to be gained by utilizing knowledge acquired from large-scale molecular studies of model organisms. These studies have informed us about the nature of many genes in a few model organisms, and how these genes are organized into complexes, pathways and different levels of biological processes. Further, these datasets have uncovered general properties of genetic interaction networks, and which of these are well-conserved among species. Unfortunately, these large datasets are often collected using gene knockout mutants, or other loss of function mutations, and these mutations tend to be deleterious. Relatively little is known about how natural mutations or beneficial mutations might differ, and we might expect these types of mutations to be categorically different in their properties and epistatic relationships. In addition, there is still relatively little known about how epistatic relationships change with changing environments, especially for mutations that were beneficial in the original environment. By considering the data produced by large-scale studies when making evolutionary hypotheses and utilizing the biological tools developed by molecular biologists to perform directed experiments when the necessary information is lacking, I believe that we can greatly improve our ability to interpret evolutionary genetic data and our predictions for evolutionary trajectories.

Chapter 2

Too much of a good thing: The unique and repeated paths toward copper adaptation

2.1 Introduction

In his book, *Wonderful Life* (GOULD 1989, p. 51), Stephen J. Gould famously opined that evolution is a historical and contingent process, so much so that "any replay of the tape would lead evolution down a pathway radically different from the road actually taken." While this is undoubtedly true when one considers the full complexity of an organism, refrains are often observed in evolution at the trait level. Repeated evolution, defined as 'the independent appearance of similar phenotypic traits in distinct evolutionary lineages' (GOMPEL and PRUD'HOMME 2009) has been documented in both ecological and clinical environments at all taxonomic levels, e.g., repeated loss of stickleback lateral plates in freshwater (SCHLUTER *et al.* 2004), ecomorphs of Anolis lizards (LOSOS 1992), the acquisition of "cystic fibrosis lung" phenotypes in *Pseudomonas aeruginosa* in cystic fibrosis patients (HUSE *et al.* 2010), to name but a few. The development of sequencing technologies has recently allowed biologists to ask whether parallel genetic changes underlie observations of parallel phenotypic change. In some cases, parallel phenotypic evolution has been attributed to parallel genotypic evolution, for example, repeated changes to cis-regulatory regions of the same gene—the pigmentation gene *yellow*—underlie changes in wing pigmentation in male Drosophila (PRUD'HOMME *et al.* 2006). At the other extreme are cases where different genetic targets underlie similar phenotypic shifts; for example, yeast adapting to rich media converged in fitness via a variety of genetic mechanisms (KRYAZHIMSKIY *et al.* 2014), and beach mice adapting to sandy coastal dunes from the Gulf and Atlantic coasts of Florida converged in coat coloration via different mutations (MANCEAU *et al.* 2010). In such cases, unique evolutionary trajectories at the genetic level appear repeatable at the phenotypic level.

The degree of phenotypic repeatability is inherently linked with the genomic target size of appropriate mutations, with single-locus Mendelian traits with fewer target sites (and hence higher repeatability) at one extreme and quantitative traits at the other extreme. Even when multiple genes underlie a selected trait, however, there may be relatively few sites that, when mutated, have the magnitude of effect and sufficiently minor deleterious side effects to improve fitness overall (STERN 2013). Such pleiotropic constraints are thought to explain why cis-regulatory sites more often contribute to adaptation than trans-regulatory changes (GOMPEL *et al.* 2005; STERN 2000). The size of the population and the manner in which it reproduces are also critical. Large populations have access to rarer mutations, particularly those of large effect (BURCH and CHAO 1999), increasing the chance that the best of these mutations will fix in independent evolutionary trials (BELL and COLLINS 2008). Mutations with particularly high fitness are also more likely to fix in asexual

populations, because clonal interference reduces the chance that minor-effect mutations establish (ROZEN *et al.* 2002), unless adaptive mutations are so common that coalitions of mutations establish together (FOGLE *et al.* 2008; LANG *et al.* 2013).

The nature and severity of environmental challenge will also affect the degree of repeatability at both the genotypic and phenotypic levels. If the environmental change is so severe that the population cannot replace itself and there are only a small fraction of mutations whose benefits are large enough to bring absolute fitness above one (BELL and COLLINS 2008), then adaptation would be more repeatable. On the other hand, if an organism is adapting via mutations whose effects are small relative to the distance to the fitness optimum, nearly half of mutations are predicted to be beneficial (FISHER 1930), and adaptation would be less repeatable. The genomic target size must also depend on the nature of mutations required: when adaptation can be accomplished by the loss of a function, adaptive mutations can potentially arise in any step along the pathway leading to that function via a variety of mechanisms (e.g., single base pair changes leading to premature stop codons early within a gene, movement of transposable elements within a gene, mutations in the promoter that alter transcription factor binding sites, etc.). In contrast, if the environmental challenge requires the appearance of a novel trait, or an alteration of an existing trait, the number of genomic targets is likely diminished. Despite these long-standing theoretical predictions, empirical data have only recently been catching up, largely due to breakthroughs in sequencing technology (e.g, SCHNEEBERGER 2014).

In this study, we set out to determine the repeatability of adaptive evolution at the genotypic and phenotypic levels using short-term experiments with the yeast, *Saccharomyces cerevisiae*. We purposefully employed a short-term experimental design in an attempt to avoid the potential influence of epistasis limiting the mutations that are sampled (CHOU *et al.* 2011; KVITEK and SHERLOCK 2011). The design of the experiment was similar to a previous study in our group where we examined adaptation to the fungicide nystatin (GERSTEIN *et al.* 2012). In both cases, multiple isogenic lines of yeast were exposed to inhibitory levels of either nystatin or copper, with levels chosen to be slightly higher than those in which growth occurred reliably. Lines that showed growth were isolated and analyzed. Through whole-genome sequencing of 35 lines that evolved tolerance in the nystatin experiment, we found that adaptation repeatedly involved the same four genes in a single pathway leading to the production of ergosterol (GERSTEIN *et al.* 2012), the membrane-bound target of nystatin (WOODS 1971). Indeed, of the 20 unique mutations identified, 18 involved the same two genes (11 different sites in *ERG3* and 7 in *ERG6*). In hindsight, the highly repeated nature of this adaptation may well be explained by the narrowness of the environmental challenge: the cells can survive by blocking the production of ergosterol, and this can be accomplished through loss-of-function mutations in the ergosterol biosynthesis pathway (particularly in *ERG3* or *ERG6*). We thus set out to assess the degree of repeatability in the face of an entirely different environmental challenge: high copper concentrations, where loss-of-function mutations are less expected.

Copper is a micro-nutrient that is essential for several different enzymatic processes in yeast (cytochrome oxidase involved in respiration, superoxide dismutase involved in defense against oxidative damage, and the Fet3p ferro-oxidase involved in iron uptake, GRADEN and WINGE 1997). Thus, unlike nystatin, cells cannot entirely block copper uptake. On the other hand, copper is extremely toxic at high concentrations, both because it displaces other metal co-factors from proteins and because it produces highly reactive oxygen

species, including damaging hydroxyl radicals (PEÑA *et al.* 1999). The fact that multiple cellular processes require copper, that multiple cellular compartments are involved in copper sequestration (especially vacuoles and mitochondria), and that multiple processes are impacted negatively by copper (PEÑA *et al.* 1999) suggests that adaptation to high copper concentrations may occur through a variety of mechanisms. Here we report the results of a short-term adaptation experiment to this toxic but essential metal. Through whole genome sequencing, we identify the nature of the genetic changes that underlay the evolutionary rescue of 34 lines of *S. cerevisiae* exposed to inhibitory copper concentrations.

2.2 Materials and Methods

2.2.1 Evolution of haploid mutation lines

Mutations were acquired in haploid lines of the common lab strain, BY4741 (*MATa his3△1 leu2△0 met15△0 ura3△0*), obtained from Open Biosystems in 2009. Preliminary experiments determined that BY4741 grown in liquid YPD + 12.5mM $CuSO_4$ does not show consistent growth, but that some populations begin growing at different times, a stochastic pattern of growth we have previously shown to be consistent with beneficial mutations in other environments (GERSTEIN *et al.* 2012). To initiate mutation acquisition, we streaked BY4741 from frozen onto a YPD plate and randomly chose a single colony to grow overnight in 10mL YPD, shaking at 30°C. We added 4.5mL of this common wild type stock to 185.5mL YPD + 12.5mM $CuSO_4$ (hereafter referred to as 'copper12 medium'). We placed 1mL aliquots into 180 inner wells of three 96 deep-well boxes, with 1mL of dH2O in the outer wells. Inner wells were used, with dH2O in the outer wells, in an effort to reduce evaporation. Boxes were maintained shaking on a platform shaker at 30°C. All boxes were checked daily by visual examination of the bottom of the wells. Growth was recorded when we saw precipitate on the bottom of a well and was first observed after 7 days of incubation (Table A.1). Twenty-four hours after growth was first seen, we manually mixed the well and froze 500μL of culture in 15% glycerol. In this way we isolated 56 'putative mutation lines' within 14 days, post-inoculation.

At the end of the mutation-accumulation phase we struck each putative mutation line from the freezer stock onto a single YPD plate and grew them at 30°C. Two of the putative mutation lines did not grow within 72 hours and were excluded from this point forward. Fourteen lines exhibited very small colonies, typical of petite colonies that have lost mitochondrial function. One of our initial goals in acquiring these mutation lines was to measure their fitness in heterozygous form; to avoid assaying non-nuclear mutations, these lines were also excluded from the set of lines we genotyped and phenotyped. From each remaining line we haphazardly chose eight colonies and inoculated each colony into 1mL copper12 medium and 1mL YPD (by using the same pipette tip) and incubated them at 30°C with shaking. In six cases, none of the eight colonies grew in copper12 medium within 72 hours, leaving us with 34 copper-adapted mutation lines (CBM: 'Copper beneficial mutation' lines, Table 2.1). From the paired YPD culture descended from the same colony (limiting exposure of our stocks to copper), 500μL was added to 500μL 30% glycerol and frozen.

Table 2.1: Mutations identified in the CBM lines. *CUP1* coverage for each line is provided in the second column and does not account for additional copies via chrVIII aneuploidy.

CBM line	*CUP1* coverage	Genome Position (chr.bp)	Gene	Mutation (Watson strand)	Position (from 5' end)	Amino acid change	Exchangeability
CBM1	1.61	X.412600	*VTC4*	C>T	800	Trp>Stop	
		XI.105507	*FAS1*	G>T	4837	Val>Phe	0.207
		XVI.420661	intergenic	A>T			
CBM2	2.00	chrII aneuploidy					
CBM3	2.48	VII.150650	intergenic	G>T			
		chrII aneuploidy					
CBM4	3.26	mito.24277[a]	*COX1*[b]	1D indel (GG C/- CC)	10460	intron	
CBM5	3.78	X.413174	*VTC4*	C>A	226	Glu>Stop	
		X.654261	intergenic[c]	T>C			
		XIV.284255	intergenic	T>G			
CBM6	3.69	III.100061	*BUD3*	G>A	3781	Gly>Arg	0.178
		IV.319466	*VAM6*	T>A	655	Lys>Stop	
		mito.59168	21S_RRNA	A>G	1160	Lys>Arg	0.440
		mito.69322	tRNA-Arg	C>G	34	Arg>Gly	0.251
CBM7	0.91	II.365359	*TRM7*	C>T	361	Val>Ile	0.537
		III.306327	intergenic	G>T			
		IV.143017	YDL176W	G>T	921	Ser>Ser	
		IV.177435	*CLB3*	T>G	663	Thr>Thr	
		V.392908	*BOI2*	C>T	805	Glu>Lys	0.323
		VII.949946	*SMI1*	C>A	954	Lys>Asn	0.457
		IX.370383	intergenic	C>G			
		XV.215888	*MAM3*	C>G	250	Gly>Arg	0.178
		chrVIII aneuploidy					
		chrXVI aneuploidy					
CBM11	2.46	X.413020	*VTC4*	1D indel (GG A/- AA)	380	Phe>Ser+frameshift	
		XI.566200	*CCP1*	A>G	999	Phe>Phe	
		XII.605283	intergenic	1D indel (TT A/- AA)			
		chrII aneuploidy					
CBM13	4.02	X.412247	*VTC4*	C>A	1153	Glu>Stop	
		X.654261	intergenic[c]	T>C			
CBM14	2.15	XV.215018	*MAM3*	C>T	1120	Val>Ile	0.537
CBM16	0.28	VII.480836	*PMA1*	A>T	1831	Phe>Ile	0.181
		chrII aneuploidy					

Continued on next page

Table 2.1 – continued from previous page

CBM line	*CUP1* coverage	Genome Position (chr.bp)	Gene	Mutation (Watson strand)	Position (from 5' end)	Amino acid change	Exchangeability
CBM17	0.98	X.412325	*VTC4*	A>G	1075	Tyr>His	0.197
		XIII.711207	*ESC1*	C>T	4075	Leu>Phe	0.336
		XIII.821262	*FCP1*	T>C	1007	Leu>Ser	0.212
		chrVIII aneuploidy					
		chrXVI aneuploidy					
CBM18	2.80	V.303094	*VTC1*	G>T	289	Asp>Tyr	0.227
		VII.548326	*GSC2*	C>T	63	Asp>Asp	
		XI.646356-onwards	*FLO10*[d]	A>G			
CBM20	1.82	VII.480463	*PMA1*[e]	G>T	2204	Ala>Asp	0.193
		XV.215332	*MAM3*	C>T	806	Ser>Asn	0.390
		XVI.84024	YPL247C	C>T	173	Gly>Asp	0.188
		chrII aneuploidy					
CBM21	1.12	VII.971165	*PFK1*	G>C	2570	Pro>Arg	0.254
		X.654261	intergenic[c]	T>C			
		chrII aneuploidy					
		chrIII aneuploidy					
		chrVIII aneuploidy					
CBM22	0.78	V.302818	*VTC1*	1D indel (CA C/- CA)	13	Pro>His+frameshift	
		chrVIII aneuploidy					
		chrXVI aneuploidy					
CBM24	0.77	IV.805485	intergenic	A>G			
		IV.805517	intergenic	G>A			
CBM25	2.28	IV.530697-onwards	*ENA5*[f]	A>G			
		IX.621992	*MLP1*	G>T	2188	Glu>Stop	
CBM26	0.66	VII.480470	*PMA1*	T>G	2197	Thr>Pro	0.164
		chrI aneuploidy					
		chrV aneuploidy					
		chrVIII aneuploidy					
CBM29	1.08	VII.1376	intergenic	A>G			
		VII.480463	*PMA1*[e]	G>T	2204	Ala>Asp	0.193
		XV.566240	intergenic	G>C			
		chrII aneuploidy					
CBM30	3.30	chrII aneuploidy					
CBM33	2.40	VII.618173	*VHT1*	G>C	1686	Ile>Met	0.279
		VIII.321332	*SBE22*	A>T	919	Met>Leu	0.513
		X.412080	*VTC4*[g]	C>T	1320	Trp>Stop	

Continued on next page

Table 2.1 – continued from previous page

CBM line	*CUP1* coverage	Genome Position (chr.bp)	Gene	Mutation (Watson strand)	Position (from 5' end)	Amino acid change	Exchangeability
		mito.24277[a]	*COX1*[b]	1D indel (GG C/- CC)	10460	intron	
CBM34	2.92	X.412080	*VTC4*[g]	C>T	1320	Trp>Stop	
		XI.364518	intergenic	complex 1I indel (GA>AAT)			
		mito.24277[a]	*COX1*[b]	1D indel (GG C/- CC)	10460	intron	
CBM36	2.15	X.412080	*VTC4*[g]	C>T	1320	Trp>Stop	
		mito.24277[a]	*COX1*[b]	1D indel (GG C/- CC)	10460	intron	
CBM37	2.04	X.412080	*VTC4*[g]	C>T	1320	Trp>Stop	
		mito.24277[a]	*COX1*[b]	1D indel (GG C/- CC)	10460	intron	
CBM44[h]	1.39	X.412080	*VTC4*[g]	C>T	1320	Trp>Stop	
		mito.24277[a]	*COX1*[b]	1D indel (GG C/- CC)	10460	intron	
CBM45	2.68	X.412080	*VTC4*[g]	C>T	1320	Trp>Stop	
		mito.24277[a]	*COX1*[b]	1D indel (GG C/- CC)	10460	intron	
CBM46	2.74	X.412643	*VTC4*	T>A	757	Arg>Stop	
		XI.438478	*DID4*	A>G	701	Gln>Arg	0.366
CBM47	2.34	V.302909	*VTC1*	C>A	104	Ser>Stop	
CBM49	3.14	V.438349	intergenic	G>C			
		XII.1034221	*HMG2*	C>T	1595	Pro>Leu	0.258
		XIII.420239	intergenic	A>C			
		XIV.265933	*GCR2*	T>A	598	Lys>Stop	
CBM51	2.50	II.444465	*FES1*	C>T	229	Asp>Asn	0.201
		IV.310552	intergenic	A>G			
CBM53	2.88	V.180433	*PRP22*	C>T	1593	Ile>Ile	
CBM54	1.98	VII.1077964	*MAL12*	G>T	1366	Gly>Stop	
CBM55	2.25	(no mutations except to *CUP1*)					

[a] This mutation falls in an intron of *COX1* but causes a frameshift in an overlapping predicted gene, A15_Beta.

[b] Identical *COX1* mutation observed in seven different lines.

[c] Identical intergenic mutation observed in three different lines.

[d] The alignment formed a 100% match to the beginning of *FLO10* until XI.647464, at which point the alignment switched to a perfect match to a similar region downstream, starting at XI.648031. *In silico* qPCR confirmed the absence of unique intervening sites (CACCAGCTCTTCCTGGTCGT and CACCAGCTCTTCCTGGTCGT) within the FASTQ files for CBM18 (but present in other CBM lines), indicating a deletion in this region.

[e] Identical *PMA1* mutation observed in two different lines.

[f] The alignment within this region exhibited a 100% match to the beginning of *ENA5* but switched to a 100% match to *ENA1* from approximately site IV.527743, suggesting a deletion. Because of the highly repetitive nature of this array, *in silico* qPCR was unable to uniquely identify the missing positions.

[g] Identical *VTC4* mutation observed in six different lines.

[h] CBM44 was sequenced from the original population, not the representative colony.

2.2.2 Sequencing of haploid mutation lines

Freezer culture from each CBM line was streaked onto YPD plates and grown for 48 hours at 30°C. We haphazardly picked a single colony for each line and grew it for 24 hours in 50mL of YPD at 30°C with shaking. Genomic DNA was extracted using standard protocols (SAMBROOK and RUSSELL 2001). Protocols supplied by Illumina were followed to create barcoded libraries for each line (2011 Illumina, Inc., all rights reserved). We sequenced 100bp single-end fragments for each line, pooling 12 uniquely barcoded strains in each lane on an Illumina HighSeq 2000. Twelve samples were rerun to obtain sufficient depth of coverage using 100bp paired-end fragments: CBM18, 20, 21, 22, 24, 25, 26, 29, 30, 34, 36, 44.

The resulting genomic sequence data was processed using Illumina's CASAVA-1.8.0 as in GERSTEIN *et al.* (2012). We called SNPs and small insertions and deletions using configure-Build.pl and parsed the output files with custom UNIX and perl scripts. We took advantage of Illumina data from the previous set of experiments with nystatin (GERSTEIN *et al.* 2012), which were initiated from the same BY4741 culture, to determine the mutations that are common to our strain background yet different from the S288C reference genome (scergenome.fasta downloaded from the Saccharomyces Genome Database, `http://downloads.yeastgenome.org/genome_release/r64/`); all such differences were removed from the dataset. Given that our lines were haploid, mutations called as heterozygous were discarded (likely alignment errors), as were SNP and indel calls of low quality ($Q < 20$). Remaining variants were checked in the alignments, using tview in samtools-0.1.7a (LI *et al.* 2009). SNPs were independently called using the bwa software package to perform the alignment along with samtools-0.1.7a to identify SNPs, using the -bq 1 option to limit data to reliable alignments (LI *et al.* 2009), confirming all SNPs found by CASAVA (Table 2.1).

To assess chromosomal aneuploidy events, the total depth of coverage for each chromosome was calculated as the proportion of sequenced sites mapping to a particular chromosome, relative to the proportion of known mapped sites located on that chromosome within the yeast reference genome (as reported by configureBuild.pl in Illumina's CASAVA-1.8.0 package).

Intergenic mutations were analyzed for gains and losses of predicted TF binding sites using Cis-BP, a tool offered by the online Catalog of Direct and Inferred Sequence Binding Preferences (available at `http://cisbp.ccbr.utoronto.ca/TFTools.php`). Cis-BP compares two sequences (i.e., one wildtype and one mutant allele) for differential transcription factor binding inferred based on the relationship between similarity in DNA binding domain amino acid sequence and DNA sequence preferences (WEIRAUCH *et al.* 2014).

2.2.3 Expected frequency of mutations causing non-synonymous and stop codons

The expected frequency of mutations that would generate a particular type of amino acid change (synonymous, non-synonymous, or stop) was calculated from the observed codon frequency in *S. cerevisiae* (`http://downloads.yeastgenome.org/unpublished_data/codon/ysc.gene.cod`; produced by J. Michael Cherry based on the 6216 ORFs within the "Saccharomyces Genome Database" (SGD) as of January 1999). For each position in each codon, the frequency of all possible mutations was calculated according

to the observed spectrum of mutations reported by LYNCH *et al.* (2008) based on previous studies in yeast. (Similar results were obtained using other mutation spectra, including a uniform distribution, the spectrum observed by LYNCH *et al.* (2008) in their mutation-accumulation study, and the observed mutation spectrum in this study.)

Summing over the whole genome, the expected frequency of mutations leading to stop codons is 5.78%. The expected frequency of non-synonymous mutations is 73.0% among all possible codon changes or 77.6% among only the synonymous and non-synonymous changes (excluding those going to or from a stop codon). The expected frequency of mutations causing any change to the amino acid sequence is 78.9%, which is similar to the expectation used previously (78.7%) based on a uniform frequency of mutations (WENGER *et al.* 2011). Because mutations are biased towards transitions and away from G/C, we recommend using the estimates reported here, which are based on the greatest amount of data regarding the mutational spectrum (LYNCH *et al.* 2008).

2.2.4 Determination of *CUP1* copy number

Using samtools-0.1.7a, the alignments of all CBM lines were manually checked at genes that are known or suspected to be important for acclimation to high levels of copper in *S. cerevisiae*, with a particular focus on genes that were previously identified to be up-regulated under high levels of copper: *BSD2*, *CCC2*, *COX23*, *CTR2*, *CUP1-1*, *CUP1-2*, *CUP2*, *FET3*, *FMP23*, *GEF1*, *HAA1*, *PCA1*, *SCO1*, *SCO2*, *SLF1*, *VMA3*. The alignments were normal for all of these genes (including 500bp up and downstream), except *CUP1-1* and *CUP1-2* on chromosome VIII. In this region, large gaps were consistently found spanning the duplicated copies of these genes, caused by alignment ambiguities in this tandem repeat region.

To measure *CUP1* copy number without having to rely on alignments, we carried out the bioinformatics equivalent of a qPCR analysis (*in silico* qPCR; GERSTEIN *et al.* 2014) by using the unix command "grep" to directly count the number of fastq fragments containing "primers" in the *CUP1* region. Specifically, we summed the number of fragments containing the 16 bp fragment from the very beginning and from the very end of *CUP1*, plus two 16 bp fragments between *CUP1-1* and *CUP1-2* (TTTCAAGAGAACATTT and GGGTGGTGAAGTAATA), searching for all four in the forward and reverse directions (e.g., using "zgrep TTTCAAGAGAACATTT *fastq*"). We then repeated this *in silico* qPCR procedure for three unique genes on chromosome VIII as controls (using the first 16 bp of *DED81*, *DUR3*, *RIX1* in both the forward and reverse direction). A BLAST search was used to confirm that these fragments aligned only to the appropriate genes (`http://www.yeastgenome.org/cgi-bin/blast-sgd.pl`). We also conducted this procedure with the 35 BMN lines isolated in nystatin ('beneficial mutation nystatin'; GERSTEIN *et al.* (2012)), which we initiated from the same ancestral genetic background, providing a baseline for comparison. Relative to the three control genes, the BMN lines had an average of 18.13 copies of *CUP1* (range: 12.40 – 30.45). Note that although the S288C reference genome on the Saccharomyces Genome Database (SGD) reports only two *CUP1* copies, an isolate of S288C was recently found to contain about 14 copies by Southern analysis (ZHAO *et al.* 2014). Our data are thus consistent with our ancestral BY4741 strain having undergone amplification in this region, and we report the number of *CUP1* copies in our copper adaptation strains relative to the

Table 2.2: Oligonucleotides employed for real time PCR (RT), Southern blot analysis (SB), and genotyping (GT) in the forward (F) and reverse (R) directions.

Primer name	Sequence	Experiment
CUP1-F	AGCTGCAAAAATAATGAACAATGC	RT
CUP1-R	GCATTTGTCGTCGCTGTTACA	RT
TAF10-F	AAGTTGTTCTGACGGTGAACGA	RT
TAF10-R	GCGACCTATATTGAGCCCGTATT	RT
CUP1-F	5Biosg/TTAATTAACTTCCAAAATGAAGGTCA	SB
CUP1-R	5Biosg/AGACTATTCGTTTCATTTCCCAGAG	SB
MAM3-F	AATGAGTGCCGATACCATCC	GT
MAM3-R	GATTCGTCCCAATCTTTTGC	GT
VTC4-F	GTTCATGATCTAGCAAAGTTTTCG	GT
VTC4-R	GGTAACCAAAATGGGATTGAA	GT
LYS2-F	TCAAGGGCTGAAAAGACAATCAA	GT
LYS2-R	CGACGCAAAGAGATGAAACCA	GT

average across the BMN lines.

To test whether levels of *CUP1* inferred from *in silico* qPCR were consistent with levels of *CUP1* transcription, we assayed RNA levels using quantitative real-time PCR (qPCR). Detailed methods are provided in Section A.1.1. Briefly, we chose 10 CBM lines that spanned the range of *CUP1* copy number. A single colony of each CBM line and two colonies of BY4741 were inoculated into 1mL YPD + 5.5mM $CuSO_4$ (a lower concentration was used to allow growth of all lines, including BY4741) and grown for 12 hours at 30°C with shaking, at which point RNA was isolated using the RNEasy Mini Kit from Qiagen, following the yeast protocol. Oligonucleotides for qPCR (Table 2.2) were designed using Primer Express (ABI). mRNA levels of *TAF10* were used for normalization, because *TAF10* exhibits stable expression across strains and conditions (TESTE *et al.* 2009).

2.2.5 Phenotypic assays of CBM lines

To determine the extent of copper tolerance acquired, we conducted dose-response experiments in deep-well boxes. Each CBM line was struck from frozen onto YPD and grown for 48 hours at 30°C. A single colony was then haphazardly chosen from each line and inoculated into 10mL YPD, shaking overnight at 30°C. The optical density of all lines was standardized to the least dense line and 200μL of standardized culture was added to 400μL YPD; 15μL was then inoculated into 1mL of 8 different levels of YPD + $CuSO_4$ (0mM, 4mM, 8mM, 9mM, 10mM, 11mM, 12mM, 14mM). Four replicates were grown for each line in each level of copper. Boxes were maintained shaking on a bench top shaker at 30°C. After 72 hours we manually mixed each well and the optical density (OD) of 200μL of culture was measured on a BioTek plate reader. With this data, we determined the IC_{50} (half-maximal inhibitory concentration) of copper using a maximum likelihood fitting procedure, as previously described (GERSTEIN *et al.* 2012).

To assess whether there was a correlation between the ability to grow in elevated levels of copper and fitness in other environments, we conducted a series of growth rate experiments using the Bioscreen C Microbiological Workstation (Thermo Labsystems) to automate OD readings. From the rise in OD, growth rates were estimated under multiple environmental conditions: YPD + 8mM $CuSO_4$ ('copper8'); YPD, a standard laboratory rich medium; YPG, a medium that requires yeast cells to respire; and YPD + iron (ferric citrate). The latter environment was of particular interest because of copper's role in iron uptake via the Fet3p ferro-oxidase, so growth was assayed at three levels of ferric citrate: 10mM, 40mM and 60mM; we only present the 40mM results in the main text because results were highly correlated across the iron concentrations (Table A.2). Copper (0.2M Cu(II)$SO_4 \cdot 5H_2O$) and iron (1M $C_6H_5FeO_7$) stocks were made in distilled water. Iron stock was made at least three days prior to use with occasional vortexing and mild heating to keep the ferric citrate in solution. In both cases, copper or iron stock was added after YPD was autoclaved, roughly one hour before the addition of yeast culture.

Each growth rate assay was initiated in a similar manner to the IC_{50} assays. Cultures from BY4741 and all CBM lines were struck from frozen and grown on YPD plates incubated at 30°C for 2-3 days. Four or five colonies from BY4741 and a single colony from each CBM line was then inoculated into 10mL YPD, shaking overnight at 30°C. Optical density from overnight culture was standardized, and a 1:101 dilution was conducted into the appropriate medium. For each line, five random wells spanning two 100-well honeycomb plates were filled with 150μL of diluted culture. Plates were incubated at 30°C with maximum shaking for 24 hours on a Bioscreen C, with automated OD readings every 30 minutes. From the raw data, we extracted the maximum growth rate using a non-parametric spline fit performed by a custom R script (GERSTEIN *et al.* 2012). The maximum growth rates from the replicates of each CBM line were statistically compared against all replicates initiated from BY4741 using a t-test (replicates involving separate wells from a single Bioscreen C plate).

2.2.6 Copper tolerance of deletion lines

To assess whether intragenic mutations that arose within our CBM lines are phenotypically similar to knockout mutations, we measured copper tolerance (IC_{50}) of 21 gene deletion lines (GIAEVER *et al.* 2002), representing all of the available knockouts for the characterized genes that had mutated in our study (excluding the uncharacterized YPL247C and YDL176W). BY4741 is the progenitor of both the deletion collection and our ancestral strain background, allowing a direct comparison of the impact of deleting these genes. Tolerance (IC_{50}) was determined as above from OD measurements taken across an array of copper concentrations at 24 hours. Tolerance assays were conducted in the Bioscreen C and replicated twice, running simultaneously on two different machines.

2.2.7 Tetrad dissections to isolate single mutations

To separate the effects of single mutations from other mutations present in the evolved lines (including extra copies of *CUP1*), we crossed all of the CBM lines with BY4739, which has a common genotype

yet opposite mating type and different auxotrophies than BY4741, the progenitor of our lines. We then attempted to sporulate the resulting diploid lines, focusing on a subset that contained each common mutation or aneuploidy and the fewest number of additional mutations (∼1/3 of the lines). Detailed methods are provided in Section A.1.2.

We encountered substantial difficulties in obtaining tetrads from our strains; BY4741, a derivative of S288c, is known to be a poor sporulator (BEN-ARI *et al.* 2006; DEUTSCHBAUER and DAVIS 2005). In particular, despite many attempts, no tetrads were obtained for CBM16 (*PMA1* mutation plus chrII aneuploidy), CBM26 (*PMA1* mutation plus chrI, chrV and chrVIII aneuploidy), CBM29 (*PMA1* mutation plus chrII aneuploidy), CBM47 (*VTC1* mutation), or CBM55 (no mutation identified other than extra copies of *CUP1*).

We were able to sporulate CBM2 (chrII aneuploidy), CBM14 (*MAM3* mutation), CBM25 (*MLP1* and *ENA5* mutations), and CBM34 (*VTC4* mutation). CBM25 was not initially chosen for tetrad dissection but was dissected as a contaminate of CBM22 (*VTC1* plus chrVIII and chrXVI aneuploidy), as detected by subsequent sequencing. CBM25 contaminating cells were likely positively selected during the sporulation procedure given that the aneuploid lines in our experiment, like CBM22, had very low sporulation rates.

The genotype of resulting spores was then determined (see Section A.1.2; PCR primer information in Table 2.2). In brief, for CBM14 and CBM34 tetrad lines, *MAM3* and *VTC4*, respectively, were amplified by PCR. All SNPs showed the expected 2:2 segregation pattern in the four spores of each dissected tetrad. CBM25 spores were sequenced on Illumina HiSeq 2000, which is when the strain was discovered to be CBM25 (bearing a mutation in *MLP1* and *ENA5*), not CBM22. The segregation pattern for the additional copy of chrII in CBM2 spores was determined by the segregation patterns of *LYS2* alleles. To quantify the segregation patterns of *CUP1* among the spores, Southern blots with *CUP1* specific probes were performed. We isolated genomic DNA and ran a Southern blot on three separate occasions for each spore. Band intensity was quantified in ImageJ (ABRAMOFF *et al.* 2004) using the “background corrected density" macro to estimate *CUP1* copy number.

2.2.8 Fitness effect of single mutations on growth rate and copper tolerance

To measure the fitness effects of the mutations isolated by tetrad dissection, growth rate assays were conducted within the Bioscreen C using either YPD + 9mM $CuSO_4$ (‘copper9’) or YPD, as described above with the following exceptions. Yeast was occasionally taken from a lawn plated from frozen cells rather than from single colonies (the sporulated lines had been bottlenecked to a single colony just prior to freezing, and so a second bottleneck at this stage was less essential). Optical density was not standardized for the copper9 experiments as this was deemed to have little effect on inferred growth rates. For each line, two (copper9) or four (YPD) non-adjacent wells were filled with 150μL of diluted culture and allowed to grow for 24 hours. This procedure was performed three times in the copper9 environment, and once in YPD to determine whether these lines were affected in their ability to grow in the nutrient-rich environment. The mean maximum growth rate was determined for each Bioscreen C assay in the copper9 environment, and statistics were performed using these means as data points.

Copper tolerance was determined for a subset of spores through dose-response experiments as described for the knock-out lines except that two separate Bioscreen C runs were performed, with two replicate wells per run (Figure A.1). Specifically, we assayed IC_{50} for two spores (among all of the tetrads for each line) that carried mutations of interest but had low *CUP1* copy number.

2.2.9 Data Accessibility

To facilitate data reuse, all genomic fastq files have been deposited in the NCBI-SRA database under the accession code PRJNA261735. The remaining raw data and statistical analyses have been deposited in the Dryad Digital Repository (doi:10.5061/dryad.5gp25).

2.3 Results

We recovered a broad spectrum of genetic changes across 34 lines exposed to initially inhibitory levels of copper (Figure 2.1, Table 2.1). Most lines contained multiple mutations, in contrast to our previous results in nystatin (GERSTEIN *et al.* 2012), which is consistent with the longer waiting period before growth was observed (4-7 days with nystatin, 7-14 days with copper). All lines except for two (CBM2, one of the five lines isolated on the first day, and CBM55, one of the eleven lines isolated on the last day) contained one or more single base-pair mutations. In total, there were 57 unique base-pair changes, including four single base-pair deletions and one single base-pair insertion that also resulted in a basepair change. Beyond changes to single sites, there were several large-scale mutations. Twelve lines exhibited chromosomal aneuploidy (Figure 2.1B), and three lines (CBM6, CBM7, CBM17) appeared to have low mtDNA coverage, outside of the range of lines from our previous study with nystatin (Figure 2.1C). In addition, two changes involved deletions within repetitive regions, one in CBM25 involving the tandem array of P-type ATPase sodium pumps (*ENA5*, *ENA2*, and *ENA1*) and the second in CBM18 involving the flocculation gene *FLO10* (see details in Table 2.1).

The most frequent mutation across all lines was copy number alteration of the *CUP1* locus. Based on *in silico* qPCR, *CUP1* estimates were, on average, 2.17 times higher than estimates from the 35 lines obtained in nystatin. *CUP1* copy number was estimated to be above the entire range of nystatin lines for 24 of the 34 CBM lines, while two lines (CBM16 and 26) both exhibited *CUP1* levels lower than the range of nystatin lines (Figure 2.1C). *CUP1* is a metallothionein protein that binds copper in *S. cerevisiae*. It is present as a tandem duplication on chrVIII in the S288C reference strain, and amplification of this locus is known to be a common mutation that confers increased resistance to copper (ADAMO *et al.* 2012; FOGEL and WELCH 1982; FOGEL *et al.* 1983). Disomy for chrVIII has also been shown to increase copper tolerance (FOGEL and WELCH 1982). Indeed, including cases with chrVIII aneuploidy, 27 out of the 34 CBM lines have increased *CUP1* copy number above the range of BMN lines (one of the exceptions, CBM24, is the least copper-tolerant of our CBM lines).

When expression level of *CUP1* was investigated by qPCR in a subset of lines and compared with the *in silico* qPCR estimates, it was found that the slope was positive and significant when forced through the point

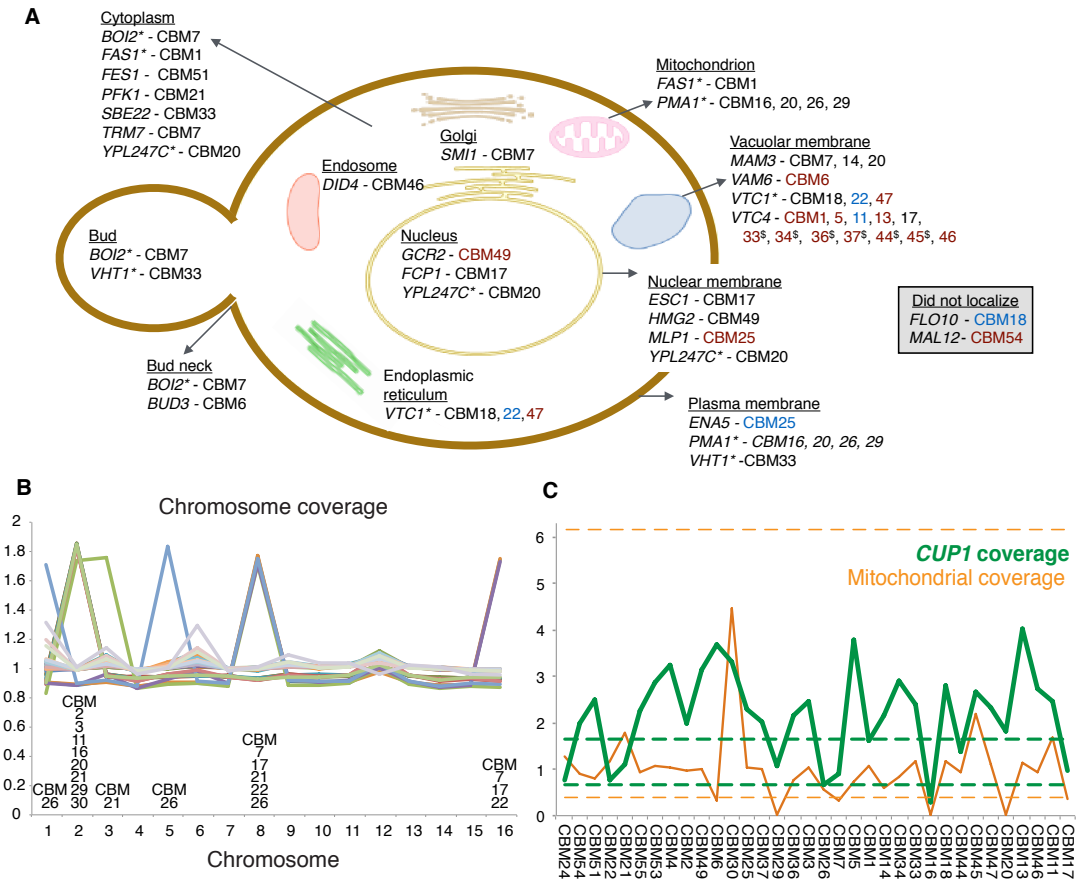

Figure 2.1: Observed mutations in copper-adaptation lines. A. Genes mutated within the CBM lines (see Table 2.1 for specific mutations) illustrated based on previous localization studies of the genes involved (references in Table 2.1). The colour of line names reflects the type of mutation; black for a nonsynonymous amino acid change, red for a premature stop codon, and blue for an indel or rearrangement (synonymous changes and RNA genes not shown). Genes that localized to more than one location are listed multiply and identified by (*). CBM lines 33, 34, 36, 37 and 45 contained the same *VTC4* mutation, indicated by ($). B. Chromosomal aneuploidy was prevalent, appearing in 11 of 34 lines. Coverage is determined by the average number of reads across the chromosome, compared to the reference strain. C. *CUP1* copy number (green line) and mitochondrial coverage (orange line) for each CBM line. *CUP1* copy number is measured relative to the average level observed in our parallel nystatin study, which showed negligible variation in copy number (range of the 35 BMN lines shown as dashed green lines). Mitochondrial depth of coverage in the CBM lines was divided by ten and is presented relative to the average depth of coverage from mapped nuclear DNA (the equivalent range for the 35 BMN lines shown as dashed orange lines). Lines are ordered according to increasing copper tolerance (Figure 2.2).

(1,1), which assumes that both axes are scaled to the ancestor (even though, technically, the derived nystatin lines and not BY4741 were used as the control in the *in silico* qPCR assays; p = 0.02, Figure A.2A). The slope was still positive, but not significant, otherwise (p = 0.27).

2.3.1 Single base-pair changes

Of the 57 unique single base-pair changes (Table 2.1), 15 were present in intergenic regions (two of which were indels), and one single base-pair deletion was present in the intron of *COX1*. The remaining 41 unique single base-pair mutations were found within 29 different genes, whose products localize to many different cellular structures (Figure 2.1A). Five of these mutations were synonymous changes, 24 were nonsynonymous changes, 10 were premature stop codons, and two were frameshift mutations caused by single base-pair deletions.

Four sites were altered in the exact same way in multiple lines (a mutation at the intergenic site X.654261 in three lines, mito.24277 within an intron of *COX1* in seven lines, VII.480463 causing an amino acid change in *PMA1* in two lines, and X.412080 causing a stop codon in *VTC4* in six lines). As discussed previously (GERSTEIN *et al.* 2012), we cannot distinguish between repeated mutational hits and either a single ancestral mutation that amplified during the growth of the strain prior to being separated into lines (i.e., during growth of the ancestral colony, followed by overnight growth in 10mL YPD) or contamination during the sampling of lines on previous days. To be conservative, we consider the same mutation in multiple lines to be non-independent and count them as having arisen only once in the statistical analyses below.

Four genes acquired multiple independent mutations, involving different positions in different strains. Fourteen lines acquired mutations in one of two subunits of the vacuolar transporter chaperone complex, *VTC1* (3 unique mutations in 3 lines) or *VTC4* (7 unique mutations in 12 lines); four lines acquired mutations in the plasma membrane H+-ATPase *PMA1* (3 unique mutations); and three lines acquired unique mutations in *MAM3*, a protein required for normal mitochondrial morphology (ENTIAN *et al.* 1999).

Given the 6607 ORFs within the *S. cerevisiae* genome (`http://www.yeastgenome.org/cache/genomeSnapshot.html`), the data are enriched for multiply hit genes. Specifically, there is a 99% chance that the 41 genic mutations would either hit different genes (first line in equation 2.1) or would hit one gene twice but no more (second line in equation 2.1):

$$\begin{aligned} 0.99 &= \prod_{i=1}^{41} \frac{6607-(i-1)}{6607} \\ &+ \sum_{j=2}^{41} \left(\prod_{i=1}^{j-1} \frac{6607-(i-1)}{6607} \right) \frac{j-1}{6607} \left(\prod_{i=j+1}^{41} \frac{6607-(i-2)}{6607} \right), \end{aligned} \tag{2.1}$$

assuming that ORFs are roughly equal in length. Thus, seeing even one gene bearing mutations at three or more independent sites is highly unlikely, and we conclude that positive selection acted upon the mutations in *VTC1*, *VTC4*, *PMA1*, and *MAM3*.

Excluding the indels, the single base-pair mutations that occurred within exons generated a stop codon

much more often than predicted by chance (10/39 = 25.6%, $p = 0.00006$, exact one-tailed binomial test with expectation of 5.78% based on the mutational spectrum in yeast, see Materials and Methods). This result remains marginally significant when we focus only on genes hit once and exclude the four multiply hit genes (4/25 = 16%, $p = 0.053$, expectation of 5.78%).

On the other hand, the fraction of unique mutations that fall within an exon rather than a non-coding region is not significantly greater than the expected fraction in *S. cerevisiae* (41/57 = 72.0%, $p = 0.63$, expectation of 72.9% from ALEXANDER *et al.* 2010). Similarly, among the synonymous and non-synonymous mutations, non-synonymous changes did not occur more often than expected (including all changes: 24/29 = 82.8%, $p = 0.34$; excluding multiply hit genes: 16/21 = 76.2%, p = 0.67; both exact one-tailed binomial tests with an expected fraction of 77.6%). Furthermore, the mean exchangeability score (an empirically-based measure of the change in protein function following a particular amino acid change, YAMPOLSKY and STOLTZFUS 2005) of our observed amino acid changes (0.294) was within one standard error of the grand mean for mutations in yeast (0.31, calculated based on the mutational spectrum reported in LYNCH *et al.* 2008). These tests are likely conservative, however, because selection against deleterious amino acid changes would have eliminated non-synonymous mutations from our dataset, making it difficult to detect an enrichment of amino acid changes due to positive selection.

The set of genes whose protein products were altered is not enriched for either a specific GO term or a particular pathway (based on YeastMine analysis, BALAKRISHNAN *et al.* 2012), although a significant number of mutated genes localize to the plasma membrane (*PMA1*, *ENA5*, and *VHT1*), the nuclear membrane (*ESC1*, *HMG2*, *MLP1*, and *YPL247C*), and the vacuolar membrane (*MAM3*, *VAM6*, *VTC1*, and *VTC4*). The set of genes is also enriched for three of the MIPS functional classification groups: vacuole or lysosome (*VAM6*, *VTC1*, and *VTC4*), cation transport (*ENA5*, *PMA1*, and *VTC1*), and protein synthesis (*TRM7* and *FES1*) (identified using Funspec, ROBINSON *et al.* 2002).

Of the characterized genes that bore mutations, 21 were available from the yeast knockout collection (GIAEVER *et al.* 2002). Relative to BY4741, 13 lines showed a significant increase in copper tolerance, and two showed a significant decrease in copper tolerance (Figure A.3). This assay supports the idea that a number of the singly-hit genes might contain mutations that influenced copper tolerance.

To identify potential regulatory changes caused by the 15 intergenic mutations we found, we assessed whether predicted transcription factor (TF) binding sites were gained or lost using Cis-BP (WEIRAUCH *et al.* 2014) (Table A.3). One of the positions (in *CBM1*) is not predicted to be at a TF binding site, while the remaining 14 were split among changes that caused both gains and losses (five mutations), only gains (four mutations), and only losses (five mutations). We identified a number of commonalities among the mutations, including two sets of transcription factor binding sites that were each lost together three times (one set involved members of the Forkhead family, *FKH2* and *HCM1*; and a second set involved *NHP6A*, *NHP6B*, and *PHO2*), two that were gained together three times (*GAT1* and *GLN3*, members of the GATA family), and some that were both gained and lost (particularly *ORC2* and *SUM1*). Among TF binding site mutations that were within 500bp and 5' of the start site of a gene, only one gene (*RPP1A*) was listed in the "Saccharomyces Genome Database" (SGD) as having an effect on metal tolerance, although two others affected vacuolar functioning (*MUK1* and YIR007W) and one affected mitochondrial functioning (*COX4*). We did

not, however, directly measure the effects of the intergenic mutations.

2.3.2 Aneuploidies

Chromosomal aneuploidy was common, appearing in one-third of all CBM lines (Figure 2.1B). All aneuploid lines had an extra copy of either chrII or chrVIII (one line had both). chrII aneuploidy was generally found by itself, only one of the eight lines with chrII aneuploidy carried additional aneuploid chromosomes (chrIII and chrVIII). In contrast, chrVIII aneuploidy never appeared in isolation. Three of the five cases of chrVIII aneuploidy also contained an extra copy of chrXVI and one line contained additional copies of chrI and chrV (in addition to the aforementioned line containing chrII and chrIII).

2.3.3 Mutagenic effects of copper

While selection must underlie the repeated spread of mutations affecting the genes that were multiply hit, it is possible that copper exposure directly altered the rate and nature of mutations that arose during the experiment. Indeed, exposure to high concentrations of copper is known to be mutagenic in experiments that directly expose DNA to copper (TKESHELASHVILI *et al.* 1991). There is no evidence, however, for an elevated base-pair mutation rate in our experiment. Focusing only on nucleotide changes (not indels), we observed 52 unique single base-pair changes across the 34 lines isolated over the course of 7-14 days (average 11.0 days until isolation, Table A.1). By comparison, in our previous study where the same ancestral strain was exposed to nystatin, we observed 35 mutations among 35 lines isolated over the course of 4-7 days (average 4.7 days until isolation). Thus, if anything, slightly more mutations accumulated per line per day in nystatin (0.21) than in copper (0.14), although the difference is not significant (p = 0.076, two-tailed exact binomial test with $n = 52 + 35$ mutation events and a proportion expected in copper given by 0.693 given that there were 34 lines $\times$ 11.0 days in copper and 35×4.7 in nystatin). Furthermore, while previous *in vitro* work indicates that copper should induce an excess of G:C $\rightarrow$ A:T mutations (TKESHELASHVILI *et al.* 1991), the spectrum of single base-pair mutations observed within this study (8 A:T $\rightarrow$ G:C, 14 G:C $\rightarrow$ A:T, 6 A:T $\rightarrow$ T:A, 12 G:C $\rightarrow$ T:A, 5 A:T $\rightarrow$ C:G, 7 G:C $\rightarrow$ C:G) is not significantly different from the mutational spectra for yeast reported by LYNCH *et al.* (2008) (see their Table 1), either based on prior studies (χ^2= 6.76, df = 5, p = 0.239) or based on their mutation-accumulation experiment (χ^2= 1.48, df = 5, p = 0.915) .

We did, however, observe many more aneuploid events with copper (affecting 12/34 lines) than with nystatin (affecting 1/35 lines, GERSTEIN *et al.* 2012), but this is only marginally significant if we account for the greater number of days until isolation (p = 0.11, two-tailed exact binomial test with $n = 12 + 1$ aneuploid lines, where the proportion expected in copper is 0.693). Here, we have treated multiple aneuploid chromosomes within a line as a single event, in the absence of information about their independence; if they were independent, the excess of aneuploid events in the presence of copper would be very significant (p = 0.013, two-tailed exact binomial test with $n = 19 + 1$ aneuploid chromosomes). An enrichment of aneuploid events in copper may well be due to selection for aneuploidy rather than an increased mutation rate, consistent with the frequent occurrence of additional copies of chrVIII, bearing *CUP1*. Nevertheless,

previous studies with mice have found copper to be mutagenic using a micronuclei assay that is sensitive to errors in chromosome segregation during mitosis (PRÁ *et al.* 2008). We thus consider it plausible that the high frequency of aneuploidy observed in this study may have been directly due to copper exposure.

2.3.4 Phenotypic assays of CBM lines

Copper tolerance (measured as IC_{50} in deep-well boxes grown for 72 hours) was fairly similar across the 34 copper-adapted CBM lines, ranging from 8.5mM – 11.2mM (Figure 2.2A). The date that mutations were isolated does not correlate with copper tolerance (mutations are numbered based on the date of isolation; Table A.1). The number of copies of *CUP1* inferred from *in silico* qPCR (Figure 2.1C, corrected to include chrVIII aneuploidy) does not directly correlate with copper tolerance (Figure 2.2B; r = 0.16, t_{32} = 0.92, p = 0.36), yet this is likely due to the confounding effects of the other genetic changes. For example, three of the lines with the lowest *CUP1* copy number carried mutations in *PMA1* (excluding CBM24, which had low tolerance; Figure 2.2B). To tease apart the effects of these mutations, we both statistically analyzed the tolerance data collected for all lines and physically dissected the mutations via tetrad analysis in a subsample of four CBM lines (see below). A linear model with the four multiply-hit genes as well as *CUP1* copy number and chrII aneuploidy as factors indicated that *CUP1* copy number (adjusted to include chrVIII aneuploidy) as well as the presence of a mutation in *MAM3*, *PMA1*, *VTC1*, or *VTC4* were all significant predictors of copper tolerance, while chrII aneuploidy was not (*CUP1* coverage: t_{27} = 2.58, p = 0.016; *VTC1*: t_{27} = 3.73, p = 0.0009; *PMA1*: t_{27} = 2.99, p = 0.0060; *MAM3*: t_{27} = 3.11, p = 0.0044; *VTC4*: t_{27} = 6.66, $p < 0.001$; *chrII*: t_{27} = 1.67, p = 0.11).

In addition to copper tolerance, we assayed maximum growth rates in copper, as well as YPD, YPG, and iron, using the Bioscreen C plate reader (Figure 2.3). Copper tended to be more inhibitory in the small volume plates used in the Bioscreen, so we reduced copper levels to 8mM $CuSO_4$ ('copper8'). Growth rate in copper8 (Figure 2.3A) was significantly correlated with copper tolerance (Figure 2.2A , r = 0.69, $t_{32} = 5.40$, $p < 0.0001$). All lines except CBM24 grew significantly faster in copper8 than did BY4741, the ancestor ($p < 0.05$, Table A.4). Copper tolerance did not correlate with growth in any of the three other environments examined (Figure 2.3B-D; YPD: r = -0.31, t_{32} = -1.83, p = 0.077; YPG: r = -0.06, t_{32} = -0.34, p = 0.73; iron: r = -0.09, t_{31} = -0.52, p = 0.61). No lines exhibited significantly increased growth in the rich medium, YPD, while about half had significantly decreased growth (Table A.5). There was a negative, but not significant correlation between growth in YPD and growth in copper or copper tolerance (IC_{50}). Three of the four slowest growing lines in YPD carried multiple aneuploid chromosomes (CBM17, 22, 26), but otherwise the slow growing lines spanned a range of genotypes and *CUP1* copy numbers. There was greater variation in growth rates observed in YPG and in iron, with growth in these two environments being strongly correlated (r = 0.58, t_{31} = 3.98, p = 0.0004, YPG statistical results in Table A.6, iron results in Table A.2). The fifteen CBM lines that grew significantly slower in YPG and iron included all lines with *PMA1* mutations (CBM16, 20, 26, 29), chrXVI aneuploidy (CBM7, 17 and 22), and chrII aneuploidy (CBM2, 3, 11, 16, 20, 21, 29, 30), as well as all lines that lacked mtDNA (CBM16, 20, 29).

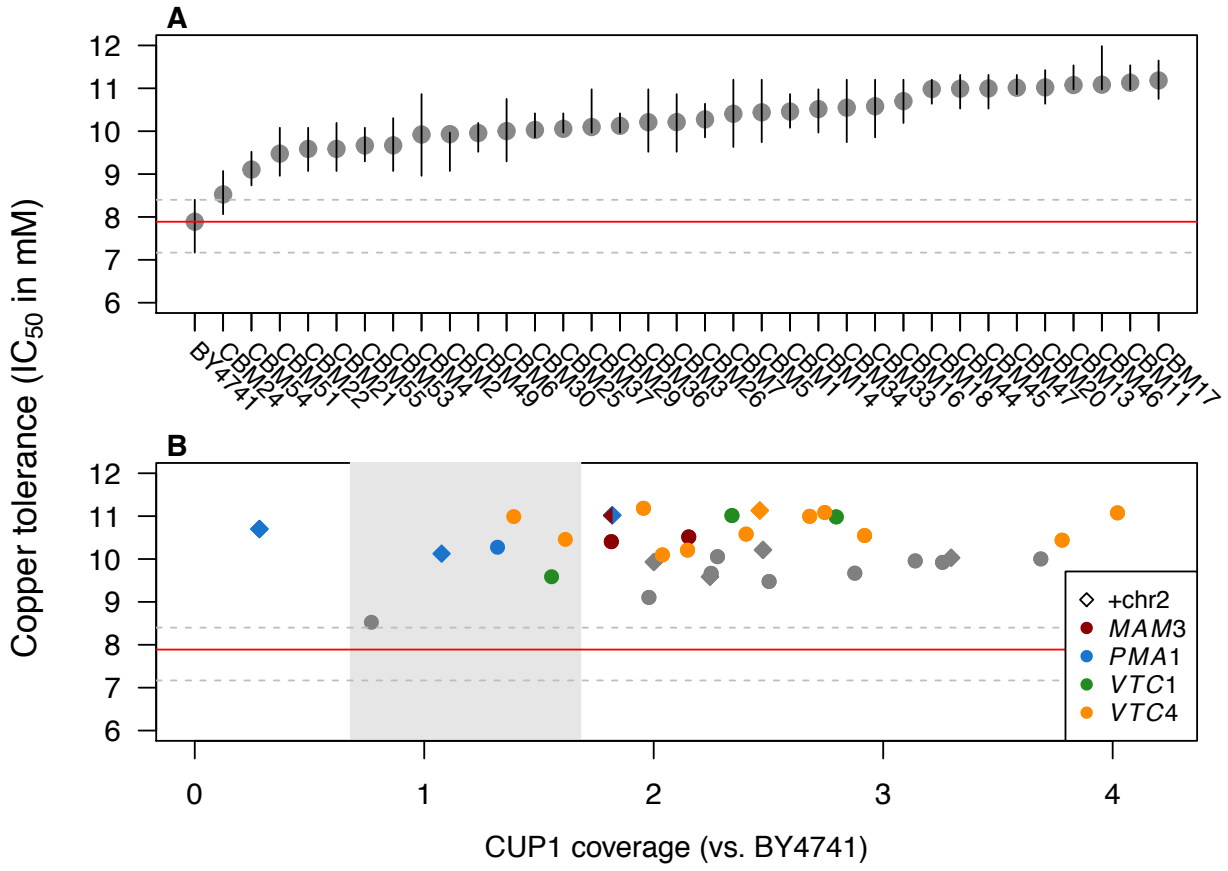

Figure 2.2: Copper tolerance across 34 copper-adapted lines ('CBM lines'). A. Lines are numbered based on the date the mutant line was isolated following exposure to copper. The order of CBM lines in other graphs is based on the order of copper tolerance depicted here, which gives the IC_{50} after 72 hours of growth in deep-well boxes (bars represent 95% confidence intervals). Tolerance of BY4741 (the ancestor) is indicated by the horizontal red line with its confidence interval indicated by dashed grey lines. B. Copper tolerance of each CBM line is generally high, regardless of the *CUP1* copy number (x-axis, accounting for duplication of chrVIII). The absence of a correlation between *CUP1* level and copper tolerance is due to the existence of additional mutations in the CBM lines, particularly in the four genes that were mutated independently (colors). Grey shading shows the range of *CUP1* observed among the BMN lines (see Methods and Figure 2.1C).

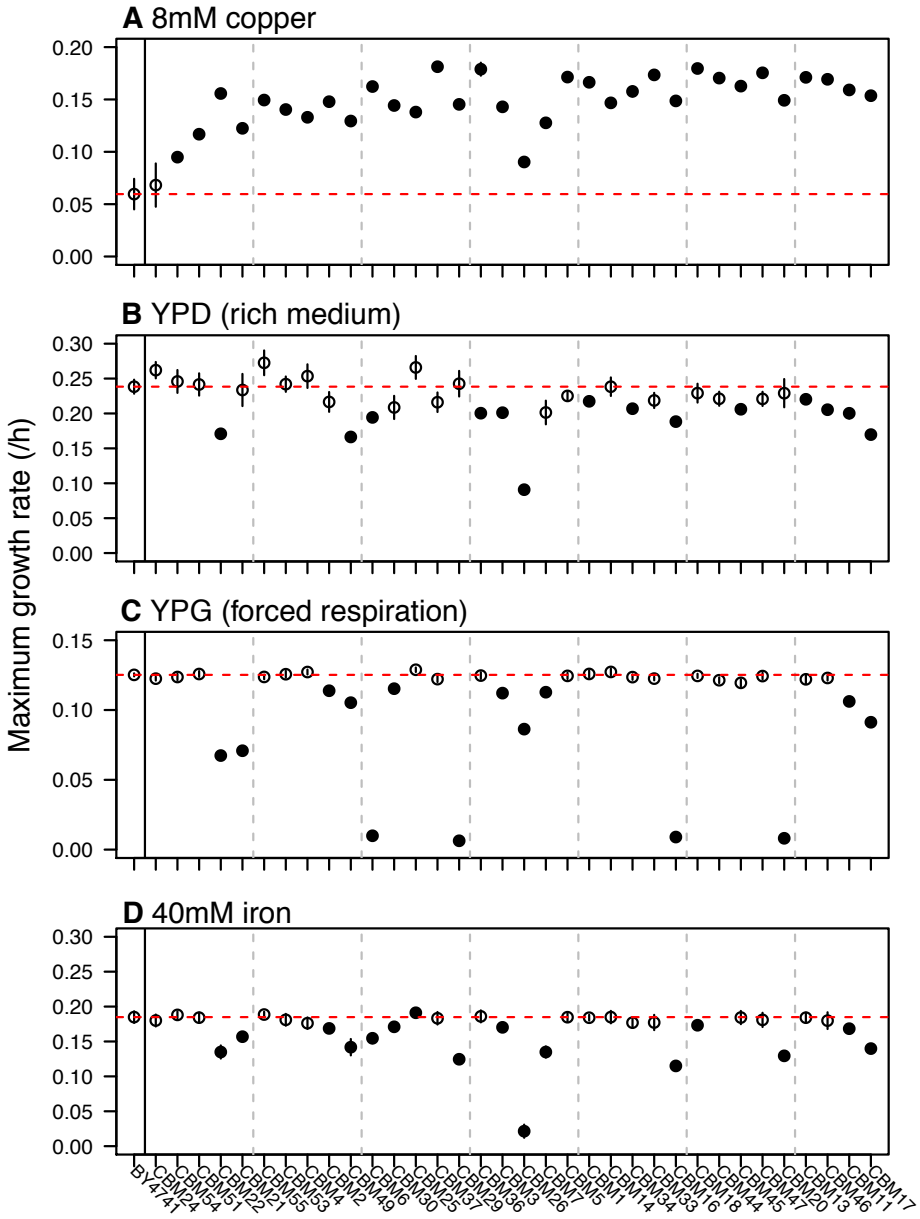

Figure 2.3: CBM lines have variable growth rates under different environmental conditions. Maximum growth rate ($\pm$1 SE) was assayed within the Bioscreen C in four different environments, each on a single day: A. YPD + 8mM $CuSO_4$, B. YPD, a standard laboratory rich medium, C. YPG, a medium that forces respiration, and D. YPD + 40 mM ferric citrate. Closed circles are lines that are significantly different from the wildtype, BY4741 (red dashed line; see Supplementary Tables A.2, A.4-A.6 for statistical information). Vertical grey dashed lines are for ease of comparison among the panels.

2.3.5 Tetrad dissections to isolate single mutations

To examine the specific effect of common mutations, we crossed CBM2 (chrII aneuploidy), CBM14 (*MAM3*), CBM25 (*MLP1* and *ENA5*), and CBM34 (*VTC4*) to BY4739 and dissected the resulting tetrads. As indicated above each panel in Figure 2.4, the four mutations and chrII aneuploidy segregated according to a 2:2 pattern ($+/-$), but segregation of *CUP1* copy number was more variable (shown by coloured circles). The variability in *CUP1* inheritance may reflect noise in the assay (band densities on Southern blots, even though measured in triplicate, Figure A.2B), or it may indicate changes in *CUP1* over the course of the tetrad line construction. Indeed, previous work has shown copy number alterations were frequent (20%) during meiosis in lines heterozygous for *CUP1* (WELCH *et al.* 1991).

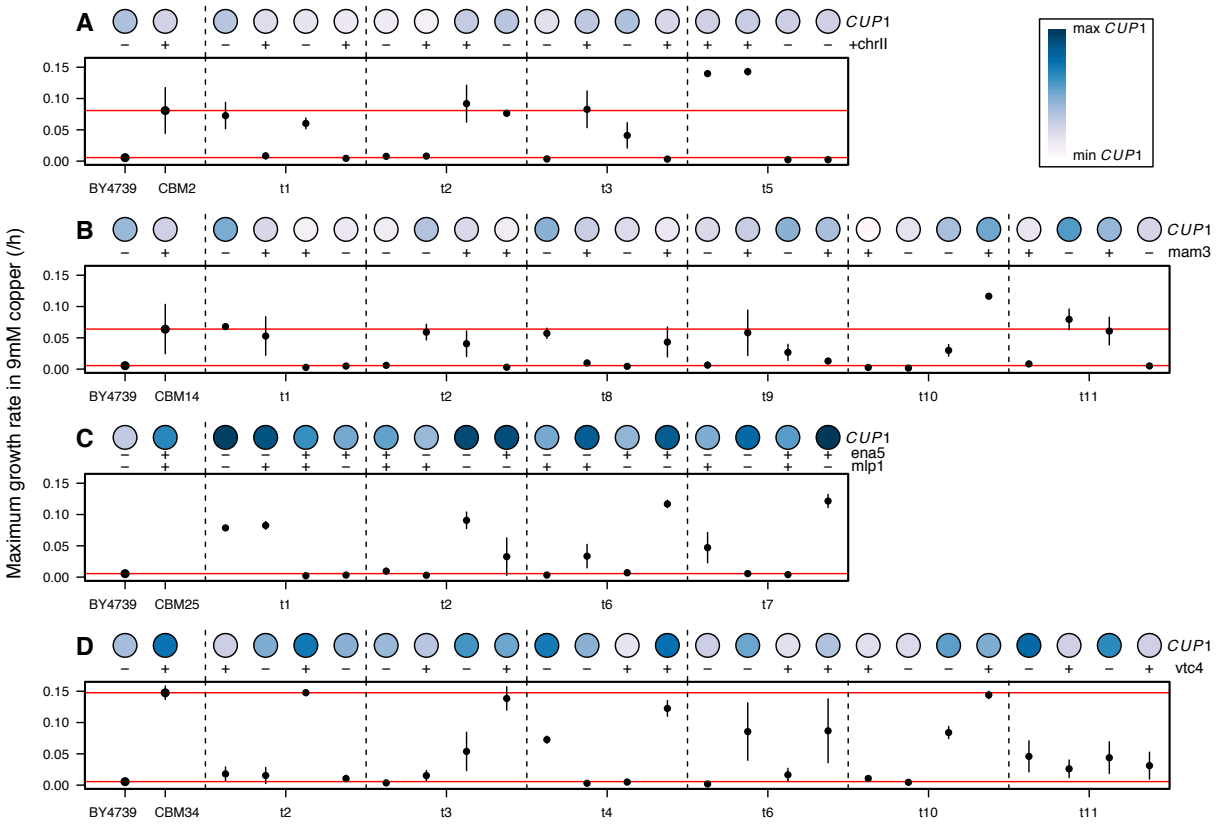

Figure 2.4: Maximum growth rates of tetrad lines in YPD+9mM $CuSO_4$. Tetrads were derived from four different CBM lines: A. CBM2, B. CBM14, C. CBM25, and D. CBM34. For each line, maximum growth rate was assayed within the Bioscreen C, with bars representing ± 1 SE across three days. The darkness of the circle represents the line's relative number of copies of *CUP1*, as assayed by Southern blot. Presence ($+$) or absence ($-$) of a segregating mutation is also noted. The maximum growth rate is shown for all tetrad lines, as well as for their two parents, BY4739, and the relevant CBM parent (red lines), except for the tetrads derived from CBM25 for which parental growth rate was not assayed (due to its initially being considered CBM22, see Materials and Methods).

Maximum growth rate in copper9 exhibited considerable variation among different spores from the same tetrad and rarely followed a strict 2:2 pattern (Figure 2.4), as expected when multiple mutations contribute to

copper tolerance. The effect of each mutation on growth in copper is more easily seen in Figure 2.5, which shows linear fits through all of the tetrad data for a given CBM line (using the mean maximum growth rate across all replicates as a single data point for each spore). The interaction terms (seen as a difference in slope) were not significant, except for a marginally significant interaction between *VTC4* and *CUP1* (t_{20} = 2.044, p = 0.054, panel D), and thus interactions were excluded from the main statistical analyses. For CBM14 and CBM34, the presence of a mutation in *MAM3* and *VTC4* (respectively), as well as *CUP1* copy number, had significant positive effects on growth rate (CBM14, *CUP1*: $t_{21} = 6.16$, $p < 0.0001$, *MAM3*: $t_{21} = 2.27$, p = 0.034; CBM34, *CUP1*: $t_{21} = 6.16$, $p < 0.0001$, *VTC4*: $t_{21} = 4.41$, p = 0.0002). By contrast, only *CUP1* copy number had a significant effect on growth rate for CBM2 and CBM25 (CBM2, *CUP1*: $t_{13} = 2.66$, p = 0.020, +chrII: $t_{13} = 1.73$, p = 0.11; CBM25, *CUP1*: $t_{12} = 3.37$, p = 0.0056, *MLP1*: $t_{12} = -0.30$, p = 0.77, *ENA5*: $t_{12} = -0.018$, p = 0.99).

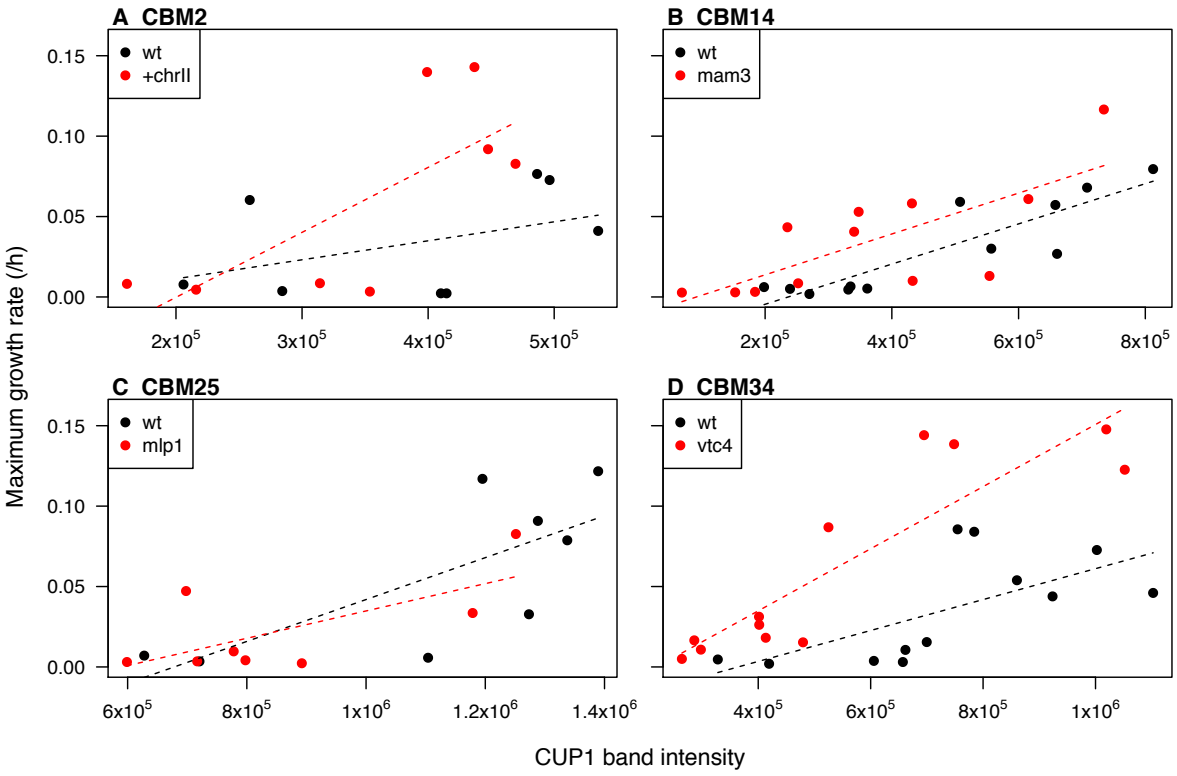

Figure 2.5: Impact of mutations on growth of tetrad lines in YPD+9mM $CuSO_4$. Dots indicate maximum growth rates for tetrad lines carrying (red) or lacking (black) the mutation of interest: A. chrII aneuploidy in CBM2, B. *MAM3* in CBM14, C. *MLP1* in CBM25, and D. *VTC4* in CBM34. For each line, linear model fits were performed with maximum growth rate as the response variable and *CUP1* copy number and the allele status of the other mutation(s) as the predictors. All fits are plotted including interaction terms between *CUP1* copy number and the other gene of interest, where *MLP1* was used as the gene of interest for CBM25 in panel C (no difference is seen when using *ENA5*).

The growth rate in copper9 may reflect an overall growth impact caused by the mutations, rather than a

specific effect on growth in copper *per se*, interfering with our ability to detect improvement. To assess this possibility, we measured growth in the rich medium, YPD, and reran the models performed above, using growth rate in YPD as the response variable. The models for CBM14, CBM25 and CBM34 contained no significant terms (Table A.7). However, the model for CBM2 indicated that the presence of an extra copy of chrII contributed to a significant decrease in growth rate in YPD (Table A.7, Figure A.4). We thus reran the copper9 linear growth models for CBM2, adjusting for growth rate in YPD by taking the difference. This model indicated that chrII aneuploidy indeed has a significant positive effect on growth in copper when controlling for its negative effect on growth in YPD (Table A.8).

Together, these tetrad analyses indicate that *CUP1* has a significant impact on growth in copper, as do the mutations in *MAM3* and *VTC4*. In addition, chrII aneuploidy has a significantly more positive impact on growth in copper than expected based on its negative effect on growth in YPD.

Finally, from among all tetrads available for each line, we measured copper tolerance (IC_{50}) for two spores that carried the mutation of interest (chrII aneuploidy, *MAM3*, *MLP1*, *VTC4*) yet exhibited low *CUP1* copy number (Figure A.1). All mutants were found to have a significantly higher copper tolerance than either the BY4741 ancestor or BY4739 parent (Figure A.5), despite their low *CUP1* copy number. IC_{50} is likely to be a more sensitive assay of resistance to copper than maximum growth rate in a single copper level, suggesting that all of these mutations increase copper tolerance.

2.3.6 Reexamining the petite mutations

After the above analyses were conducted, we reexamined the lines that displayed small colonies on YPD plates (Table A.1). To confirm that they were incapable of respiration, we assayed their growth on YPG plates. While 11 lines showed no growth, three lines (CBM9, 27, and 28) formed colonies, indicating that their mitochondria were still functional. These three lines were then whole-genome sequenced using population samples (Table A.9). Line CBM9 carried a high-frequency SNP in *PMA1*, CBM27 carried a three base pair deletion in *PMA1*, and CBM28 was aneuploid for chrIII, V, and VIII (a combination not seen in any other line). Levels of *CUP1* were assayed as in Figure 2.1C and fell within the range of the BMN lines (CBM9: 0.69; CBM27: 0.83; CBM28: 1.25), except for CBM28 once we account for its extra copy of chrVIII. Altogether, these lines provide additional confirmation of the role of *PMA1* and chrVIII aneuploidy in the evolution of copper tolerance, although we infer that the mutational routes taken by these three lines were particularly deleterious in the absence of copper, given their small colony size on YPD plates.

2.4 Discussion

Depending on the different possible directions in which the environment may change, how rapidly can evolution happen and via how many different possible pathways? The advent of rapid sequencing technology has led to an increase in studies examining the repeatability of evolution at the level of the genotype, finding that the same genes often underlie parallel and convergent evolution in natural populations and during

experimental evolution studies (CONTE *et al.* 2012; MARTIN and ORGOGOZO 2013, see references within). Whether this repeatability reflects convergence over time to fitter genotypes or a limited scope of adaptive mutations along the way remains unknown. Also unknown is the relationship between the type of evolutionary challenge and the likelihood of parallel genetic changes (STERN 2013). This study aimed to contribute to our understanding of how the type of environmental challenge influences the genomic target size of the mutations selected during the very first steps of adaptation. We used the same mutation acquisition protocol as in our previous study on nystatin resistance (GERSTEIN *et al.* 2012) to obtain mutations in the presence of an inhibitory level of copper. We predicted that a broader range of genetic solutions would underlie copper adaptation, in contrast to the nystatin study that identified a narrow genetic solution (all 35 lines had mutations within four genes in the same pathway). Previous studies have suggested that xenobiotic environments (such as antimicrobial drugs) select for repeated genetic solutions (MARTIN and ORGOGOZO 2013). By contrast, copper is a very different kind of environmental stressor—it is essential for several different enzymatic processes in yeast (GRADEN and WINGE 1997) and therefore cannot be blocked entirely from entering the cell. Copper is, however, toxic at high concentrations (PEÑA *et al.* 1999), and thus its concentration in the cell must be held in a delicate balance. As predicted, we identified a large number of mutations among our copper adaptation lines, with the level of genetic parallelism highly dependent on the type of mutation under investigation.

Increased copy number of the *CUP1* locus through tandem duplication or aneuploidy of chrVIII was by far the most common mutation, seen in 27 of the 34 copper adaptation lines. *CUP1* exists as a tandem repeat in the S288C genome, and adaptation under copper stress has previously been shown to select for amplification of this locus (ADAMO *et al.* 2012; FOGEL and WELCH 1982). Indeed, COVO *et al.* (2014) demonstrated that moving *CUP1* onto other chromosomes can efficiently select for disomy under copper stress. The number of *CUP1* copies varies among naturally-isolated wild and vineyard strains (between 1 and 18 copies among 14 wild strains, ZHAO *et al.* 2014, and 4 and 18 copies among 15 Italian vineyard strains, STROOBANTS *et al.* 2008). As *CUP1* is present in the ancestral genome as a tandem repeat, it is likely prone to alterations in copy number as a consequence of unequal crossover, gene conversion, or single-strand annealing (ZHANG *et al.* 2013). Furthermore, *CUP1* amplification seems to incur few pleiotropic costs, as seen by the lack of an observed effect of *CUP1* copy number on growth rate in YPD among our tetrad lines (Table A.7).

Chromosomal aneuploidy also repeatedly arose within our lines. Twelve of the 34 lines were aneuploid, and each of these 12 lines contained either chrVIII aneuploidy (five lines) or chrII aneuploidy (eight lines). Chromosomal aneuploidy seems to be a common route to adaptation for fungal species reproducing asexually in a diverse array of environmental stressors such as drug resistance (SELMECKI *et al.* 2009; SIONOV *et al.* 2010), high temperature (YONA *et al.* 2012), and salt (DHAR *et al.* 2011). Aneuploidy is an intriguing beneficial mutation, as it has the potential to affect many genes simultaneously, yet has a much higher reversion rate than other types of mutations. Whether such a high degree of aneuploidy would have been seen had our strains evolved for longer remains an open question. It may be that aneuploidy serves primarily as a stop-gap adaptation until other beneficial mutations appear in the genome with fewer costs (YONA *et al.* 2012). In our experiment, chrVIII aneuploidy may have been positively selected for its effect on *CUP1*

copy number, as seen by FOGEL and WELCH (1982). Similarly, chrII aneuploidy had a more beneficial effect in copper (Figure 2.5, Table A.8) than expected based on its low growth in YPD (Table A.7), perhaps because it amplified genes contributing to copper tolerance, such as *SCO1* and *SCO2* (non-adjacent gene duplicates on chrII), which function in the delivery of copper to cytochrome c oxidase in the mitochondrial inner membrane. ChrII aneuploidy was also repeatedly observed by COVO *et al.* (2014) when investigating chromosomes gained in response to copper stress. Whether the other aneuploid chromosomes had an effect on copper tolerance remains unknown. Repeated attempts to sporulate CBM22 (aneuploid for chrVIII and chrXVI) and CBM26 (aneuploid for chrI, chrV, and chrVIII) failed to yield tetrads.

The major genetic contributors to copper tolerance besides *CUP1* were four genes involved in maintaining plasma membrane potential (*PMA1*), vacuolar transport (*VTC1*, *VTC4*), and mitochondrial morphology (*MAM3*). These genes each bore several independent protein-coding mutations in different lines, which is highly unlikely in the absence of selection. Indeed, of the seven lines that did not exhibit *CUP1* amplification (relative to the range of BMN lines), six involved mutations in these genes (three in *PMA1*, one in *VTC1*, two in *VTC4*), with the remaining low-*CUP1* line exhibiting little copper tolerance (CBM24).

The types of mutations observed in *VTC4* and *VTC1* point to selection for loss of function in these genes, with several mutations inducing stop codons or frame shifts (Table 2.1). Furthermore, deletion of *VTC1* and *VTC4*, as well as *MAM3*, significantly increases copper tolerance (Figure A.3). By contrast, *PMA1* codes for a plasma membrane H+-ATPase that regulates cytoplasmic pH, and yeast are inviable when this gene is deleted. High levels of copper have previously been shown to have a deleterious effect on plasma membrane organization (FERNANDES *et al.* 2000), while strongly stimulating plasma membrane ATPase activity (FERNANDES and SÁ-CORREIA 2001). It thus seems likely that the mutations identified in *PMA1* alter, rather than inactivate, this protein, suggesting a gain (or fine-tuning) of function. Consistent with this view, none of the *PMA1* mutations involved stop codons or frame shifts (neither the four listed in Table 2.1, nor the two additional *PMA1* mutations among the lines initially categorized as petites, Table A.9).

In addition, 25 genes bore a single mutation in a single line in our experiment (Table 2.1). These unique mutations were no more likely to be nonsynonymous than expected based on the mutational spectrum and no more likely to occur within exons than expected. That said, several of the unique mutations caused stop codons (4/25), a marginally significant excess. Furthermore, 12/18 of the deletion lines tested for this subset of genes were found to have significantly altered copper tolerance when compared to BY4741 (Figure A.3), suggesting that at least some of the uniquely hit genes may be playing a role in copper tolerance. Alternatively, many of the mutations in singly-hit genes may have been neutral but spread via hitchhiking, a pervasive phenomenon in other batch culture experiments (LANG *et al.* 2013).

Beyond genetic changes, epigenetic changes may have contributed to copper adaptation. This possibility was not formally investigated in our study. We can, however, conclude that epigenetic change was not the primary cause of adaptation, given that plausible causative mutations, involving either *CUP1* or the four multiply-hit genes, occurred in every line except one (Table 2.1, Figure 2.2B). The exception, CBM24, was the least copper tolerant line and remains a possible candidate for epigenetic adaptation, although our genomic analysis may not have found all rearrangements or changes in hard-to-align regions.

One question of interest is why our copper-adapted lines carried so many more mutations, on average,

than the nystatin-adapted lines studied previously (GERSTEIN *et al.* 2012). In no case did we see a BMN line that carried two mutations thought to be adaptive; all lines carried one and only one mutation in an *ERG* gene. By contrast, 21 out of the 34 CBM lines carried more than one mutation for which we have evidence of a beneficial effect in copper (including the four multiply-hit genes, chrII aneuploidy, and *CUP1* levels above the BMN range when including chrVIII aneuploidy; 15 out of 34 lines if we exclude chrII aneuploidy). The greater contribution of multiple beneficial changes to copper adaptation cannot be explained by a larger pool of large-effect beneficial mutations or a higher mutation rate, because it took longer to observe growth in copper (7-14 days) than in nystatin (4-7 days). One possibility is that many of the adaptive mutations may not have been beneficial enough on their own to generate detectable growth; instead, they may have allowed a line to persist for longer or to expand slightly in population size, facilitating the appearance of subsequent large-effect mutations. An alternative possibility is that positive epistasis among mutations more strongly favoured the spread of secondary mutations. Our tetrad analysis provides some evidence for this possibility. Figure 2.5 shows that the effect of *VTC4* on growth rate rises with *CUP1* copy number among tetrads of CBM34, a marginally significant positive interaction ($t_{20} = 2.044$, $p = 0.054$). Similarly, the benefits of chrII aneuploidy also appear mild at low *CUP1* levels and rise with increasing *CUP1* copy number among CBM2 tetrads, although this interaction is not significant ($t_{12} = 1.61$, $p = 0.13$). In accordance with either of these explanations (mutations facilitating subsequent adaptation or positive epistasis), *VTC4* mutations occurred more often among the CBM lines with higher *CUP1* copy numbers (Table 2.1, Figure 2.2B).

This study provides an in depth analysis of how a eukaryotic organism, like yeast, takes its first few evolutionary steps towards tolerating an inhibitory but essential element, copper. Our genome-wide analysis of 34 strains found that adaptation often involved a common step (especially amplification of *CUP1*), but that routes less taken were also available. These alternate routes often involved chromosomal aneuploidy of chrII or chrVIII and four genes with roles in a wide variety of cellular functions – vacuolar transporters, mitochondrial morphology, and cytoplasmic pH regulation. Compared to our previous study of nystatin resistance (GERSTEIN *et al.* 2012), the variety of genes allowing growth in copper suggests that altered environments with more widespread effects on a cell may also provide a broader genetic basis for evolutionary recovery. Previous longer-term experimental evolution studies with microbes (e.g., CONRAD *et al.* 2009; HERRON and DOEBELI 2013; KRYAZHIMSKIY *et al.* 2014; LANG *et al.* 2013; MILLER *et al.* 2011; TENAILLON *et al.* 2012; WONG *et al.* 2012) have found a high degree of parallel evolution at the gene level, and our results suggest that this can be the case even among the very first mutations selected. Whether adaptive mutations are likely to recur depends, in theory, on the effect size of the mutations as well as their mutation rate. In our case, amplification of *CUP1* is both a relatively large effect mutation and easily acquired, contributing to its highly repeated nature, but it was also aided by the beneficial effects of many other less repeated mutations. In short, adaptation to copper is both more and less repeatable than adaptation to nystatin, with adaptation via *CUP1* representing the route most commonly taken, but with mutations affecting a variety of other cellular processes providing a diversity of less travelled paths toward copper adaptation.

Chapter 3

Widespread Genetic Incompatibilities Between First-Step Mutations During Parallel Adaptation of *Saccharomyces cerevisiae* to a Common Environment

3.1 Introduction

The number of different evolutionary pathways available to populations adapting to a new environment depends on the range and characteristics of possible genetic solutions. Even populations adapting to the same environmental challenge can diverge genetically from each other if different mutations happen to establish. The long-term impact of this initial divergence depends on the fitness interactions between the available alleles that underlie adaptation to a given environment ("epistasis"). Epistasis can run the gamut from alleles that interact positively and augment each others' fitness ("positive epistasis") to those that have negative effects on fitness in the presence of each other ("sign epistasis" WEINREICH *et al.* 2005) (Fig 3.1).

3.1.1 Epistasis and its role in evolution

The nature of epistasis is critical to broad-scale evolutionary phenomena. If all possible alleles have the same effect in all genetic backgrounds, we might expect populations that diverge initially to converge to a similar genotype and/or phenotype over time at the fitness optimum. In contrast, if some alleles are beneficial only in certain backgrounds, early genetic changes will limit future genetic options, and populations may diverge genotypically and phenotypically. Thus, the shape and 'ruggedness' of the fitness landscape is directly determined by the prevalence of sign epistasis (DE VISSER *et al.* 2011; POELWIJK *et al.* 2007, 2011).

The type of epistasis can also shape the rate of adaptation. In the case of positive epistasis, when early mutations increase the beneficial fitness effects of subsequent mutations, adaptive evolution can accelerate over time. In contrast, when epistasis is negative, i.e., when first-step mutations reduce or oppose the advantage of subsequent mutations, evolution will decelerate. The deceleration of adaptation over time has been previously found in a number of experimental evolution studies (CHOU *et al.* 2011; KHAN *et al.* 2011; KRYAZHIMSKIY *et al.* 2014; SCHENK *et al.* 2013).

Even the formation of new species rests upon epistasis between alleles present in different nascent species. A major driver of postzygotic reproductive isolation between species is the build up of Bateson-

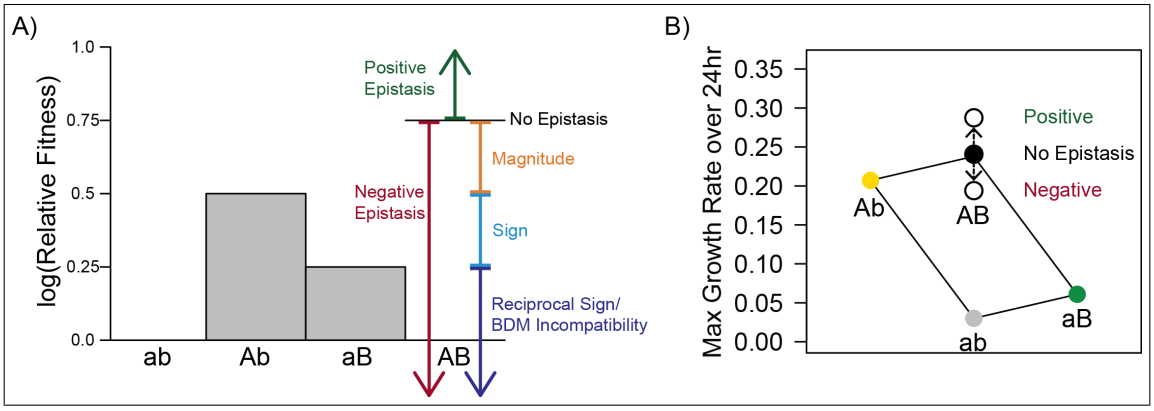

Figure 3.1: Types of epistatic relationships between mutations. (a) The type of epistasis is observed as the fitness of the single beneficial mutations (A and B) relative to the double mutant (AB). No epistasis occurs when log fitness effects are additive, as shown here (growth rate, our primary fitness measure, is calculated on a log scale). (b) Example plot showing the method used in this paper to illustrate epistatic relationships. The y-axis gives maximum growth rate over 24 hours. Point colours indicate strain genotype, where the double mutant is black, the ancestor is grey, and each single mutant has a unique colour. Lines are drawn between genotypes that are a single mutational step apart. Without epistasis, the lines form a parallelogram. Epistasis is observed as a double mutant with increased fitness (positive epistasis, higher hollow circle) or decreased fitness (negative epistasis, lower hollow circle).

Dobzhansky-Muller (BDM) genetic incompatibilities. These incompatibilities represent reciprocal sign epistasis, where alleles that work well together within a species perform poorly when combined with alleles from the other species in a hybrid individual, leading to hybrid inviability or sterility (COYNE and ORR 2004). Sign epistasis between hybrids and their more fit parental population can also contribute to speciation by reducing gene flow in one direction. With enough such asymmetric barriers acting in opposite directions, gene flow may cease entirely between populations.

All models of speciation agree that sign epistasis, and particularly reciprocal sign epistasis, is important for speciation, but they differ on why species carry different alleles. Among the models of speciation by natural selection, the classic explanation, proposed by DARWIN (1859), is that populations diverge into species because they experience different environments and so adapt in ways that often do not work well together. Because of the focus on environmental differences, this explanation has become known as "ecological speciation" (SCHLUTER 2009). A contrasting hypothesis, known as "mutation-order speciation" (SCHLUTER 2009), focuses on the chance order in which mutations arise and spread in different populations when facing the same selective environment. Even if the mutational steps that have occurred in each population are independently beneficial, combining mutations across populations need not be.

3.1.2 Determinants of epistasis

The specifics of the selective environment(s) likely have a major influence on the nature of epistasis between beneficial mutations. In environments where adaptation can occur via the elimination of a single biosynthetic pathway, complete loss-of-function mutations at one step in the pathway may lead mutations in downstream

genes to become irrelevant to fitness. Indeed, BATESON (1909) originally coined the term 'epistasis' in 1909 to describe this type of interaction, in which the action of one gene was blocked by that of another, and this is primarily how molecular geneticists continue to define the word (AVERY and WASSERMAN 1992). Considering instead partial loss-of-function mutations, genotypes combining multiple mutations may be more fit than single mutations if flow through the biosynthetic pathway is reduced by each additional mutation. In either case, we would expect double mutants to have equal or greater fitness than single mutants if knocking out a pathway is beneficial (as long as there are no pleiotropic effects beyond the pathway), and consequently sign epistasis and reproductive isolation should not arise.

On the other hand, if an intermediate phenotype is optimal in a particular environment, mutations that are beneficial on their own may overshoot the optimum when combined, causing a reduction in fitness. In this type of environment, theoretical work predicts that sign epistasis should be particularly frequent between independently selected mutations that have relatively large effects on the phenotype (FRAÏSSE *et al.* 2016).

There is also increasing evidence that epistasis is more often negative for mutations in functionally-related genes. In a large-scale screen for genetic interactions where mutations in most of the 6000 genes in the yeast *Saccharomyces cerevisiae* were tested pairwise in 23 million double mutants (including mutations in both non-essential and essential genes, although excluding ~1000 genes), COSTANZO *et al.* (2016) found that combinations of genes involved in the same biological process were enriched for negative interactions. This enrichment suggests, counter to intuition, that strongly negative fitness interactions, of the form that give rise to reproductive incompatibilities, may be more likely to accumulate between populations experiencing the same selective environment compared to those experiencing different environments.

3.1.3 Reproductive incompatibilities in nature and in the lab

To date, few incompatibilities between or within species have been genetically characterized, although recent advances in genomic sequencing technology have greatly aided the discovery of the genetic basis of speciation. For natural populations, the majority of incompatible alleles ('speciation genes') that have been characterized are found between species adapted to different local environments, presumably representing cases of ecological selection (documented in NOSIL and SCHLUTER 2011 Tables S1 and S2). For example, the build-up of a suite of plant-specific traits has allowed one species of *Drosophila* to utilize a different, normally toxic, host plant (MATSUO *et al.* 2007), and selection on soils of different salinity has caused the accumulation of QTL associated with salt tolerance in a hybrid species of *Helianthus* sunflowers beyond what is found in its parental species (LEXER *et al.* 2004). In other cases, genetic incompatibilities between natural populations have been identified where there is no clear connection to the external selective environment, including BDMs caused by the reciprocal silencing of alternative duplicate gene copies (BIKARD *et al.* 2009) or the differential accumulation of selfish genes and suppressors (see examples in MAHESHWARI and BARBASH 2011). The exact history of selection is unknown in natural populations, thus it is difficult to know whether these cases represent mutation-order or ecological selection. Natural populations of yeast also show environment-specific genetic incompatibility (including one characterized two-locus BDM HOU *et al.* 2015) though, as in other taxa, we have no knowledge of the evolutionary history that led to these interactions.

Experimental evolution studies allow direct control over the form of environmental selection, and sign

epistasis has been found in some studies that combined mutations from populations adapted to both different and similar selective environments. DETTMAN *et al.* (2008) evolved different populations of *Neurospora crassa* to high salinity and low temperature. When the evolved strains were mated, lineages adapted to different environments exhibited reduced reproductive success relative to matings between lineages adapted to the same environment, and this reduction was consistent with the action of BDM incompatibilities. A parallel study that examined populations of *S. cerevisiae* evolved to high-salinity and low-glucose for 500 generations found very similar results (DETTMAN *et al.* 2007). Follow-up work identified a BDM incompatibility between an allele of *PMA1* (a proton efflux pump) that arose under high salt adaptation and an allele of *MKT1* (a global regulator of mRNAs encoding mitochondrial proteins) that evolved in low glucose (ANDERSON *et al.* 2010). This was the first reported BDM interaction among known genes isolated from experimentally evolved strains, to our knowledge.

Sign epistasis has also been documented when combining mutations between experimentally-evolved populations adapting to the same environment. KVITEK and SHERLOCK (2011) investigated populations of asexually-propagated haploid *S. cerevisiae* evolved under glucose limitation in continuous culture for 448 generations (KAO and SHERLOCK 2008). Mutations in two genes, *MTH1* and *HXT6/HXT7*, appeared several times in independent lineages during the experiment, but never together. These mutations were shown to be individually beneficial, but they had lower competitive fitness when combined in a double mutant than either single mutant or the ancestor, showing reciprocal sign epistasis (KVITEK and SHERLOCK 2011). Negative epistasis was also prevalent among five additional strains constructed to bear two adaptive mutations that arose in different lineages, with significant negative epistasis in four out of the five comparisons, including one example of sign epistasis (KVITEK and SHERLOCK 2011). CHOU *et al.* (2014) similarly investigated epistasis using an engineered strain of *Methylobacterium extorquens* with a modified central metabolism that was dependent on a foreign pathway artificially introduced on a plasmid. These bacteria were evolved for 900 generations under conditions that utilized this pathway. All adaptive mutations decreased expression of the introduced pathway. Combining mutations, the authors found that expression levels were well predicted by the independent effects of each mutation but that expression mapped nonlinearly onto fitness, leading to sign epistasis in many cases. Collectively, these experiments demonstrate that BDMs can arise rapidly in experimental evolution studies, either when populations experience different or similar selective pressures, providing support for both ecological and mutation-order speciation.

3.1.4 Investigation of epistasis between first-step mutations

What remains unknown from long-term experiments evolved under the same selective pressure is how frequently early adaptive mutations could contribute to reproductive isolation. This raises the question of whether mutation-order speciation occurs because of incompatibilities among mutations that would be beneficial in either population or because the fixation of different initial mutations alters the subsequent selective environment experienced in different populations (i.e. divergent selection due to differences in genetic background).

We investigate, for the first time, fitness interactions among all pairwise combinations of genes bearing first-step adaptive mutations to a common selective environment. Specifically, we measured epistasis

between beneficial mutations acquired in the yeast *Saccharomyces cerevisiae* grown in the presence of the fungicide nystatin (GERSTEIN *et al.* 2012). Briefly, GERSTEIN *et al.* (2012) isolated 35 first-step mutations in 4 μM nystatin, performed genome-wide sequencing, and found that all strains carried a single mutation in one of four genes in the ergosterol biosynthesis pathway (Fig 3.2; genomic analysis revealed either no or only one other mutation present in the strains used herein, details below). We focused on one mutation in each gene and investigated the fitnesses of all six pairwise double mutants between these four mutations.

For two of these genes (*ERG6* (SGD ID: S000004467) and *ERG3* (SGD ID: S000004046)), many of the mutations found by GERSTEIN *et al.* (2012) were consistent with a complete loss of function (e.g., early stop codons, similar sterol phenotype to the whole gene knockout). The mutations occurring in the most upstream (*ERG7* (SGD ID: S000001114)) and downstream (*ERG5* (SGD ID: S000004617)) genes in the pathway, however, were not (GERSTEIN *et al.* 2012). The *erg7* mutation is a nonsynonymous change close to the end of the gene, and deletion of *ERG7* is inviable. The *erg5* mutation is an in-frame deletion and is unlikely to be a null mutation because the full gene deletion is respiratory deficient (MERZ and WESTERMANN 2009), which is not observed for this mutant (BMN35 in GERSTEIN *et al.* 2012). Thus, we also assessed whether upstream mutations in the biosynthetic pathway generally mask the effects of downstream mutations or if masking is limited to complete loss-of-function mutations.

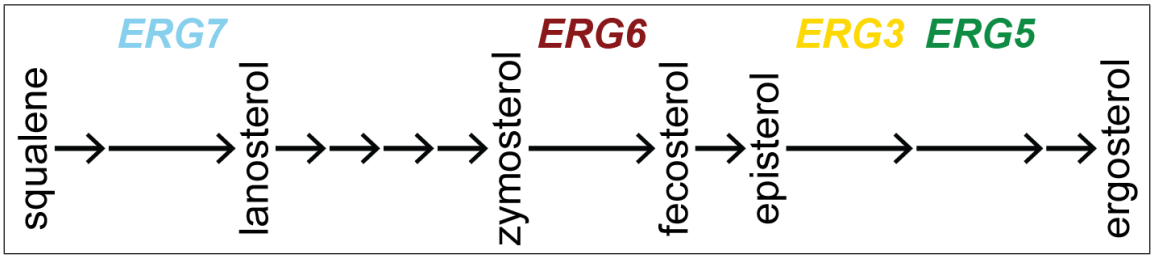

Figure 3.2: An abbreviated version of the ergosterol biosynthesis pathway. For each gene used in this study, we highlight its position in the ergosterol pathway, with gene names coloured according to the scheme used in subsequent figures. Pathway adapted from LEES *et al.* (1995).

Overall, we found that strong negative epistasis, of the type that causes some degree of reproductive isolation, between strains fixed for different mutations was surprisingly common among these first-step mutations. Indeed, the interactions were so negative that they reversed the direction of effect in over half of the double mutants, causing beneficial mutations to become deleterious when in combination and double mutants to be less fit than at least one of the two single mutants (sign epistasis) (Fig 3.1). Furthermore, in one-third of the comparisons, the double mutants were less fit than both single mutants (reciprocal sign epistasis). We assayed mutational effects in both haploid and diploid backgrounds, finding similar results regardless of ploidy, indicating that these epistatic relationships are likely to hold across stages of the yeast life cycle. Epistatic relationships for fitness were not well predicted by sterol profiles or pathway position of the mutants, however, suggesting that selection does not simply act via flux through the pathway to ergosterol.

Finally, we investigated epistasis in different concentrations of nystatin to determine how epistatic relationships, and therefore reproductive isolation, might change under different levels of environmental stress. Previous work with antibiotic resistance in bacteria has shown that the shape of fitness landscapes can be

strongly dependent on antibiotic concentrations (MIRA *et al.* 2015). Interestingly, we found that the negative interactions observed between beneficial mutations at lower concentrations of nystatin reversed sign and became increasingly positive at higher concentrations of nystatin. Indeed, only the double mutants exhibited substantial growth in the higher concentrations of nystatin tested. Thus, while combining single-step mutations generally reduced fitness in the historical nystatin environment, these same combinations were more likely than the individual mutations to allow colonization of even harsher environments.

3.2 Materials and Methods

3.2.1 Strain construction

We assayed all pairwise interactions in both haploids and diploids between four beneficial mutations acquired in the fungicide nystatin, one in each of *ERG3, ERG5, ERG6* and *ERG7* (Table 3.1). Each mutation was initially isolated in the BY4741 haploid background (*MATa his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0*, derived from S288C) and struck down to a single colony to remove standing variation. Mutations were detected by whole genome sequencing on an Illumina HighSeq 2000, followed by alignments to the S288C reference strain (GERSTEIN *et al.* 2012); few other mutations were detected besides those in the *ERG* genes. For the strains used here, only the strain containing the mutation in *ERG7* also carried a secondary mutation (in *DSC2* (SGD ID: S000005434)), whose presence or absence did not substantially alter the presented results (see details in Section B.1.2). For a complete description of the isolation of these initial strains, see GERSTEIN *et al.* (2012). All possible haploid and diploid genotypes for each pair of ERG genes were created via mating and sporulation. A brief overview of strain construction will be given here but for a detailed description, see Section B.1.1.

Table 3.1: Beneficial mutations in the strains used for the study of epistasis in the presence of nystatin (GERSTEIN *et al.* 2012).

Strain	Gene	Genome Position (Chr.Bp)	Position in Gene (nt)	Mutation	Amino Acid Change
BMN1	*ERG7*	VIII.241194	2097	C->G	Phe699Leu
	DSC2	XV.193885	916	G->A	Asp306Asn[a]
BMN9	*ERG6*	XIII.252612	379	G->C	Gly127Arg
BMN32	*ERG3*	XII.254758	898	G->C	Gly300Arg
BMN35	*ERG5*	XIII.302174 - 302233	253 - 312	60-bp deletion	

[a] Not known to affect fitness. Encodes a multi-transmembrane subunit of the DSC ubiquitin ligase complex (RYAN *et al.* 2012; TONG *et al.* 2014). Null mutant has decreased competitive fitness (BRESLOW *et al.* 2008) and decreased resistance to glycolaldehyde (JAYAKODY *et al.* 2011).

To create singly heterozygous strains, each original single mutant strain was mated to BY4739 (*MAT*α *leu2*Δ*0 lys2*Δ*0 ura3*Δ*0*) (Open Biosystems), which is isogenic with BY4741 except for the auxotrophies. *MAT*α single mutant strains were isolated by sporulation of the heterozygous diploids followed by dissection

and testing of the resulting tetrads. Throughout strain construction, histidine and lysine auxotrophies were consistently kept with the same mating types so that all haploid strains were either *MATa his3Δ1* or *MATα lys2Δ0*. Plates lacking methionine did not efficiently select against the *met15Δ0* mutation carried by the original single mutant strains, suggesting a weak effect of this mutation, and the methionine auxotrophy was not tracked.

The *MATα* single mutant strains were then mated to the original *MATa* single mutant strains to create strains that were either homozygous for one mutation or heterozygous for two mutations. The haploid double mutant strains were created through sporulation and dissection of the doubly heterozygous strains. All haploid double mutant strains were confirmed by Sanger sequencing.

We failed to obtain the *MATa erg5 erg6* double mutant haploid strain through crossing and sporulation because the two genes are linked (within 48 kb and flanking the centromere of chr XIII). For this strain, a transformation was performed by electroporation using a protocol based on CREGG (2007) to insert the mutation within *ERG6* into the *MATa erg5* genetic background; this insertion was then checked by Sanger sequencing.

Strains with one heterozygous and one homozygous mutant locus as well as double homozygous mutant strains were created by mating the *MATa* single mutant and double mutant strains to the *MATα* double mutant strains.

A representative of the diploid ancestral strain was created by mating BY4741 and BY4739.

3.2.2 Growth rate assays

We conducted a set of growth rate (fitness) assays under nystatin stress and in rich medium (YPD). The experimental design sought to ensure that data was gathered for each combination of wildtype and mutant strains across batches performed on different days. Specifically, within a batch, for a given pair of mutations in haploids and for each mating type, each ancestral strain and each single mutant was assayed twice, while each double mutant was assayed four times (the double mutant was assayed more often because it was the only genotype unique to that pair of mutations). For each pair of mutations in diploids, all possible combinations of the two genes in both heterozygous and homozygous forms (including the non-mutant) were present twice within a batch.

We measured growth in YPD and YPD + 2 μM nystatin ('nystatin2') using the Bioscreen C Microbiological Workstation (Thermo Labsystems), which measures OD in 100-well honeycomb plates. Nystatin2 was used to assay fitness because previous studies with these mutants found that 2 μM nystatin inhibits the growth of the ancestral strains while also allowing the growth of all mutant strains (GERSTEIN 2013). OD was measured automatically using the wideband filter at 30 minute intervals for 24 hours from cultures growing at 30°C with maximum continuous shaking. Longer assays were avoided because mutations and loss of heterozygosity events begin to accumulate (GERSTEIN *et al.* 2014). The maximum growth rate over 24 hours was determined by the spline with the highest slope from a loess fit through natural log transformed OD data, using a custom R script written by Richard Fitzjohn (R CORE TEAM 2015) (see ONO *et al.* 2016 for code).

For complete details on how strains were initially grown from frozen and standardized ("pre-assays")

before measuring growth ("assays"), see Section B.1.3. Briefly, each yeast replicate was grown from frozen in YPD + 0.5 μM nystatin in 100-well honeycomb plates for 72 hours in the pre-assays, unless very poor growth of the strain required otherwise, and OD was then determined. YPD + 0.5 μM nystatin was used to help prevent reversion of strains with severe growth defects in YPD and was not found to affect subsequent measures of growth, compared to a pre-assay in YPD (the first pre-assay was conducted in YPD, see details in Section B.1.3). For the main assays, honeycomb plate wells were filled with 148.5 μL of YPD or nystatin2. The yeast were then transferred from the pre-assay plates into one well each of YPD and of nystatin2, with the volume transferred determined by the maximum pre-assay OD reading (the minimum volume transferred was 1.5 μL while the maximum was 7.5 μL). Note that these transfers decreased the concentration of nystatin in the individual wells, but never by more than 0.1 μM. Strains were randomized within plates using the same map for the pre-assays and assays in a given batch.

There were not equivalent numbers of replicates for all strains after omitting some data due to low growth (if the volume to be transferred to the assay plate exceeded 7.5 μL), lack of growth, mechanical error, or because some strains had to be re-run (for details, see Table B.2). Nevertheless, at least two replicates per day on at least two days were measured for all strains in each medium (with the exception of *erg5/erg5 erg6/erg6* for which 14 replicates were all run on a single day, Table B.2; for exact numbers and days the replicates were run on, see ONO *et al.* 2016). Although the different numbers of replicates led some crosses to have less power than others, the cross with the least amount of data (*erg6* by *erg7*) was also the one where the double mutant was particularly unfit, which contributed to the difficulties in assaying fitness but also meant that epistasis was readily detected. In all cases, data for each double mutant was collected simultaneously with data on the ancestor and single mutants, allowing day effects to be factored out in the analysis.

3.2.3 Tolerance across a range of nystatin

Growth at different concentrations of nystatin was assessed following similar procedures to the growth rate assays. To prepare the strains for tolerance assays, pre-assays were again conducted to standardize initial cell concentrations. Stocks were first grown from frozen in four 96-well plates filled with 198 μL of YPD + 0.5 μM nystatin and inoculated with 2 μL of frozen culture. Strains were distributed among the four plates so that there was one replicate of the entire balanced design per plate, randomized within plate. In order to fit all strains on a single plate, some strains were excluded (*MATa erg5 erg6* and *MATa erg3 erg5*). These strains were chosen because initial assays indicated that these double mutants most closely resembled the stronger (non-*erg5*) single mutant. The plates were covered with aluminum lids and incubated at 30°C with continuous shaking at 200 rpm in a container with wet paper towels to minimize evaporative water loss. Prior to removal of the aluminum lid, plates were always spun for 1 min at 3700 rpm to ensure that all liquid was collected at the bottom.

After 72 hours, all wells were manually mixed and OD was measured on a BioTek plate reader at 630 nm. The well with the minimum OD value among the four pre-assay plates was identified and used to calculate the amount of YPD to add to each pre-assay well to standardize cell density across cultures. Wells containing only medium, those containing *erg6/erg6 erg7/erg7* (see below) and one well that appeared not to have been

inoculated were excluded from standardization. 2 μL from each well was used to inoculate the assay plates. Assay plates were prepared with 198 μL of YPD + 0, 1, 2, 4, 8, 16, 32, 64, 128, and 256 μM nystatin, with four plates per concentration. The assay plates were covered with aluminum lids and incubated at 30°C in containers with wet paper towels, shaking at 150 rpm.

Exceptions to the pre-assay protocol had to be made for strains with slower growth. 10 ml of 0.5 μM nystatin was inoculated with 15 μL of *erg6/erg6 erg7/erg7* from frozen two days before all other strains were inoculated, allowing additional growth time for this unfit strain. On the day that all other strains were inoculated from frozen, the *erg6/erg6 erg7/erg7* culture was concentrated into ∼900 μL (although growth was not observable), and 200 μL of this culture was used to replace the medium from the appropriate wells in the pre-assay plates. In addition, *erg6 erg7* (both *MATa* and *MAT*α) and *erg6/erg6* were inoculated with 2.67 μL of frozen culture (as opposed to the 2 μL used for all other strains) to compensate for their lower growth rate from frozen.

Twenty-four hours after inoculation, the aluminum lids were removed, wells were manually mixed, and the OD of each assay plate was read on a BioTek plate reader at 630 nm. Some wells had lost volume due to cracks that had developed in the plates, and these wells were omitted from analysis. Prior to analysis, the OD of the medium itself was subtracted from the final OD measurements.

3.2.4 Sterol Assay

To determine whether the sterol profiles of the single mutants, along with their position within the ergosterol pathway, predict the sterol profiles of the double mutants and whether differences in sterol profiles predict differences in fitness, a spectrophotometry-based assay was used to compare the sterol profiles of the ancestral, mutant and double mutant *MATa* strains. Sterols were extracted using the alcoholic potassium hydroxide method (ARTHINGTON-SKAGGS *et al.* 1999), as previously performed on the single mutation strains (GERSTEIN *et al.* 2012). *MATa* strains were struck from frozen onto YPD plates and grown for 65 hours. Three colonies for each strain were inoculated into two separate tubes filled with 10 mL of YPD (total of 20 mL per replicate) and incubated at 30°C on a rotor for 48 hours.

After growth, cells were harvested by centrifugation at 2700 rpm for 5 minutes, combining culture from the two tubes by performing two successive spins. The pellets were washed twice with sterile distilled water. 1.2 mL of 25% alcoholic potassium hydroxide was added to each pellet, and the tubes were vortexed for 1 minute. The tubes were then incubated in an 80°C water bath for 1 hour. After cooling the samples to room temperature, 0.4 mL of sterile distilled water and 1.2 mL of n-heptane were added to each sample, and the tubes were vortexed for 3 minutes. Samples were collected by taking 220 μL of the heptane layer and adding it to 880 μL of 95% ethanol in a 1.5 mL tube. These tubes were stored at -20°C for two days before reading the absorbance every 3 nm between 200 and 300 nm in a quartz microcuvette using a Thermo BioMate 3 spectrophotometer. Due to *a posteriori* observations that different heptane/ethanol mixtures led to different peak heights near 220 nm, we chose to use one replicate of the *erg6 erg7* strain that showed no evidence of growth (suggesting an inoculation failure), but was otherwise identically treated, as a control for standardization. As a result, only two replicates of *erg6 erg7* are presented.

3.2.5 Outlier detection and removal

Outliers in microbial fitness assays often represent either contamination by a different strain or evolution over the course of the fitness assay. In order to prevent these events from having undue influence on our analyses, we detected outliers for maximum growth rate after omitting some wells due to lack of growth and mechanical error (see details in Section B.1.3). For outlier detection, we first normalized for plate within each day. We did so by finding the global mean maximal growth rate for all ancestral strains over all days and calculating the difference between this and the mean of all ancestral strains on a given plate, yielding a plate correction value. This correction value was added to each strain from the corresponding plate. Outliers were detected by performing a two-sided Grubbs test, allowing us to detect a maximum of one outlier per strain and medium, using the R package *outliers* and the method *grubbs.test* (KOMSTA 2011; R CORE TEAM 2015). A total of eight replicates in nystatin2 and six replicates in YPD were marked as outliers and removed from all presented statistical and graphical analyses.

All qualitative relationships between strains and the main statistical conclusions were insensitive to the exclusion or inclusion of the identified outliers, with two main exceptions for the haploids in nystatin2 (see Fig B.6 and Fig B.7 for versions of Fig 3.3 and Fig 3.4 that include all outliers). These exceptions are noted in the Results and described in detail in Section B.1.4.

3.2.6 Statistical analyses

Epistasis for maximum growth rate was assessed with mixed-effects models run on either all haploid or all diploid strains together, including the genotype at each gene, their pairwise interactions, and mating type (for the haploids) as fixed effects and plate within day as a random effect, fit using restricted maximum likelihood with the *lmer* function from the *lme4* package in R (BATES *et al.* 2015; R CORE TEAM 2015). For diploids, the models were first run using only strains that were homozygous (either mutant or ancestral) for comparison to the haploid data. Significance of interaction terms (and mating type) was determined by performing an ANOVA between the full model and a model dropping that term using the *anova* function in R and fitting models using maximum likelihood.

To determine the type of epistasis present for each pair of genes, the package *lsmeans* (LENTH 2016) was used to both determine the least-squares mean for each strain in the model and to make comparisons between strains using the *contrast* function. The type of epistasis was determined by comparing the double mutant to each single mutant and each single mutant to the ancestor, and only these planned comparisons were performed. The *P*-value was adjusted for the number of tests performed using the multivariate *t* distribution (*mvt* method) in *lsmeans*. To be conservative, we based our categorization of epistasis solely on statistically significant differences. For example, if the double mutant had a lower growth rate than both single mutants but this difference was only significant in one of the two cases, it was considered an example of sign epistasis (significantly lower than one single mutant but not the other) rather than reciprocal sign epistasis.

A similar procedure was then undertaken including heterozygous diploid strains. A model was run using the *lmer* function including all diploid strains together, with plate within day as a random effect. Least-squares means were determined for all diploid genotypes from this model, and comparisons were performed between each diploid genotype and all other diploid genotypes that were one mutational step away. The

double heterozygous strains were compared to all other strains for that pair of genes because the potential progeny of the double heterozygote includes all possible genotypes and these comparisons are therefore of biological interest.

For the tolerance assay assessed across a range of concentrations of nystatin, we performed Welch's t-tests of OD after 24 hours between each double mutant and its single mutant parents (day effects were not estimated as all measurements were gathered on the same day). Because we were focused on the changing nature of epistasis, rather than any particular pairwise comparison, a correction for multiple comparisons was not performed.

Data and analyses deposited in the Dryad repository: `http://dx.doi.org/10.5061/dryad.vs370` ONO *et al.* (2016).

3.3 Results

3.3.1 Epistasis of haploids in nystatin

We characterized the epistatic interactions between pairs of mutations that act in the ergosterol biosynthesis pathway and individually confer increased fitness when exposed to the antifungal drug nystatin. Maximum growth rate of ancestral, single mutant, and double mutant genotypes was characterized in haploid strains of both mating types in YPD + 2 μM nystatin ('nystatin2'). Outlier data points were detected statistically and removed from further analyses, although we note where inclusion of outliers would have affected the results (for further details, see Section 3.2.5). The effect of mating type (and its associated auxotrophy) was not significant ($P = 0.19$), and the data for the two haploid mating types will be considered together, except where noted (see ONO *et al.* 2016 for additional statistical methods and results).

Using a mixed-effects model, all main effects of individual mutations were positive, confirming that the mutations improved growth in nystatin (Table 3.2). Double mutants were never significantly more fit than the best of the single mutants (top right panels in Fig 3.3), and all pairwise interactions exhibited significant negative (antagonistic) epistasis (Table 3.2). To assess epistasis, least-square means of maximum growth rates were inferred from the model and compared between double and single mutants and between single mutants and ancestral strains, correcting for multiple comparisons. The double mutant was significantly less fit than the fittest single mutant in four cases ("sign epistasis": *erg3 erg5*, *erg3 erg6*, *erg3 erg7* and *erg6 erg7*) and significantly less fit than both single mutants in two cases ("reciprocal sign epistasis": *erg3 erg6* and *erg6 erg7*, Table 3.2, Fig 3.3). The results are similar when fitness is measured by optical density after 24 hours of growth instead of maximum growth rate over 24 hours (Fig B.1). The strong negative interactions indicate that these alleles, when combined, confer genetic incompatibilities between the strains.

3.3.2 Comparison of epistasis between haploids and diploids

We characterized epistatic interactions of maximum growth rate for the homozygous diploid strains in nystatin2 and compared them to the haploid results to determine whether the interactions were ploidy-dependent. As in haploids, single mutations generally improved the growth of diploid homozygotes in nystatin2, although the *erg5* mutation did not do so significantly in a pairwise comparison with the ancestral strain

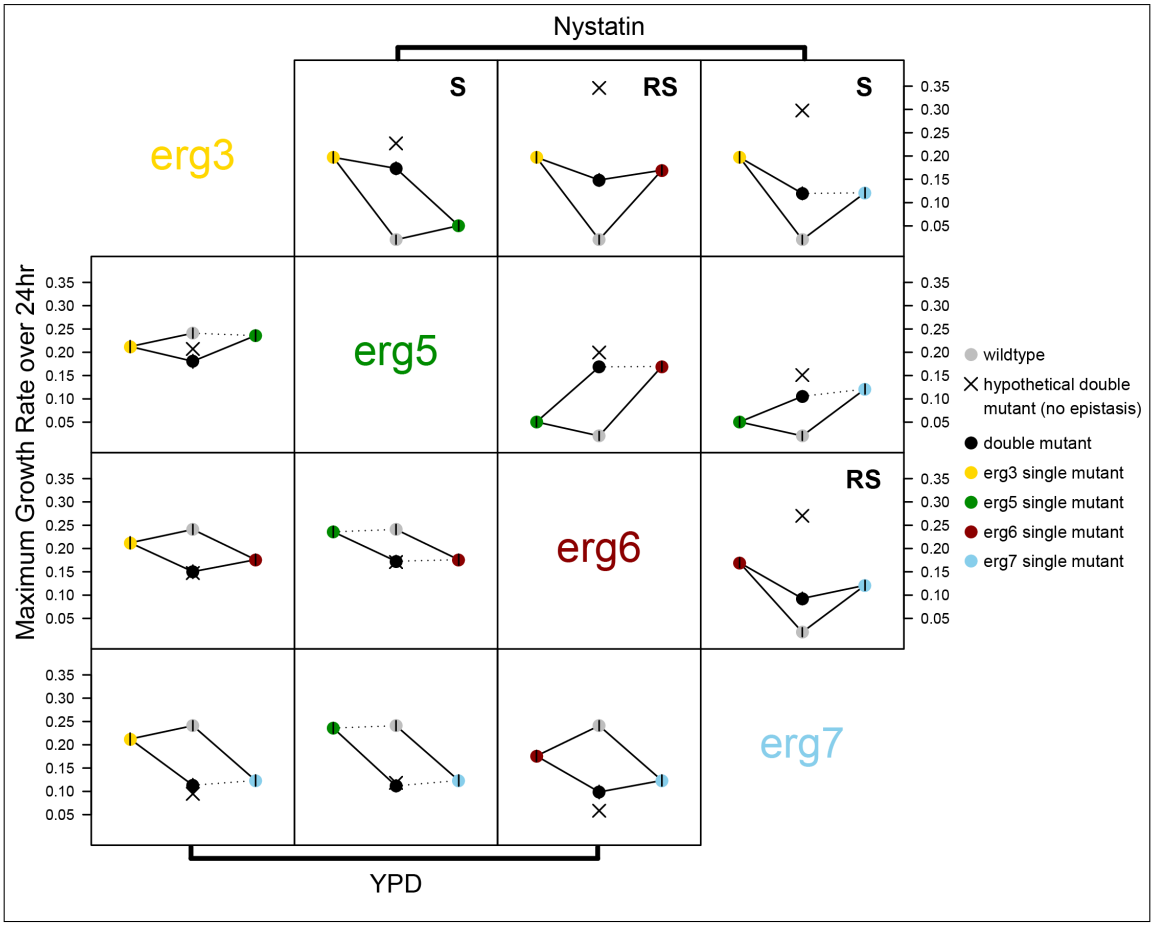

Figure 3.3: Maximum growth rate of haploid strains in nystatin2 (above diagonal) and YPD (below diagonal). Points are the fitted least-squares means of the maximum growth rates, determined in the mixed-effects model. ×'s denote the additive fitness null expectation for the double mutant, i.e., with no epistasis. Each single mutant is coloured differently, the double mutant is black, and the ancestor is grey. Vertical bars represent 95% confidence intervals of the fitted least-squares mean. Solid lines indicate significant contrasts between the fitted means, while dotted lines are non-significant. Combinations showing significant sign (S) and reciprocal sign (RS) epistasis are indicated by the presence of the abbreviation at the top of the panel. In nystatin2, the comparison between *erg3 erg5* and *erg3* is not significant when outliers are included, and the *erg3 erg6* vs. *erg6* comparison is only marginally significant ($P = 0.083$). In YPD, comparisons *erg3 erg6* vs. *erg6* and *erg6 erg7* vs. *erg7* are not significant when outliers are included. All underlying raw data and analyses can be found in ONO *et al.* (2016).

Table 3.2: Results from a mixed-effects model run on all genes using the haploid maximum growth rate data in nystatin2. Coefficients of main effects are the differences in mean maximum growth rate between the single mutant strains and the ancestral strain (difference between *MAT*α and *MATa* in the case of mating type). Coefficients of interaction terms are the differences in mean maximum growth rate between the double mutant strains and the sum of the two single mutant coefficients added to the ancestral value. *P*-values are the result of an ANOVA between the full model and one lacking that term; significant *P*-values are in bold. The last three columns refer to the type of epistasis present (Fig 3.1). "Epistasis" indicates a significant departure from an additive model of growth rates, which can be either negative or positive. "Sign" and "Reciprocal sign" refer to cases where the double mutant grows significantly less well than one or both single mutants, respectively.

Term	Coefficient	SE	*P*	Epistasis	Sign	Reciprocal sign
mating type	-0.0034	0.0026	0.19			
erg3	0.18	0.0057				
erg5	0.030	0.0049				
erg6	0.15	0.0049				
erg7	0.10	0.0049				
erg3*erg5	-0.054	0.0090	$\mathbf{3.1 \times 10^{-9}}$	negative	✓[a]	
erg3*erg6	-0.20	0.0090	$\mathbf{< 10^{-15}}$	negative	✓	✓[a]
erg3*erg7	-0.18	0.0090	$\mathbf{< 10^{-15}}$	negative	✓	
erg5*erg6	-0.031	0.0076	$\mathbf{4.6 \times 10^{-5}}$	negative		
erg5*erg7	-0.046	0.0078	$\mathbf{5.1 \times 10^{-9}}$	negative		
erg6*erg7	-0.18	0.0083	$\mathbf{< 10^{-15}}$	negative	✓	✓

[a]Not significant when outliers are included.

(Fig 3.4). Qualitatively, epistatic interactions were also similar to the haploids (Table 3.3, Fig 3.4), whether fitness was measured by maximum growth rate or optical density after 24 hours of growth (Fig B.2).

When we categorized the type of epistasis statistically for maximum growth rate, most interactions were of the same type (sign epistasis: *erg3 erg5*; reciprocal sign epistasis: *erg3 erg6* and *erg6 erg7*; negative epistasis: *erg5 erg7*). There were, however, several quantitative differences. The *erg6 erg7* double mutant was so unfit in diploids that we were often not able to standardize it properly in the growth assays (low growth, as measured by optical density, was observed in all concentrations of nystatin tested, Fig B.3). Furthermore, in two cases, epistasis was qualitatively similar, but the differences were no longer statistically significant (sign epistasis: *erg3 erg7*; negative epistasis: *erg5 erg6*).

To visualize the full diploid fitness landscape, we repeated the analysis including all heterozygous strains (open symbols in Fig 3.4, pairwise comparisons in Fig B.4). Low F_1 hybrid fitness was typical; double heterozygous strains (open diamonds) were uniformly low in fitness when compared to the homozygous single mutants (not significantly so when compared with the weak *erg5/erg5* mutant). Mutations were generally partially to fully recessive and did not have a large effect on fitness when comparing heterozygotes to wild-type at a gene, either when the other gene was wildtype (open triangles) or homozygous mutant (open circles).

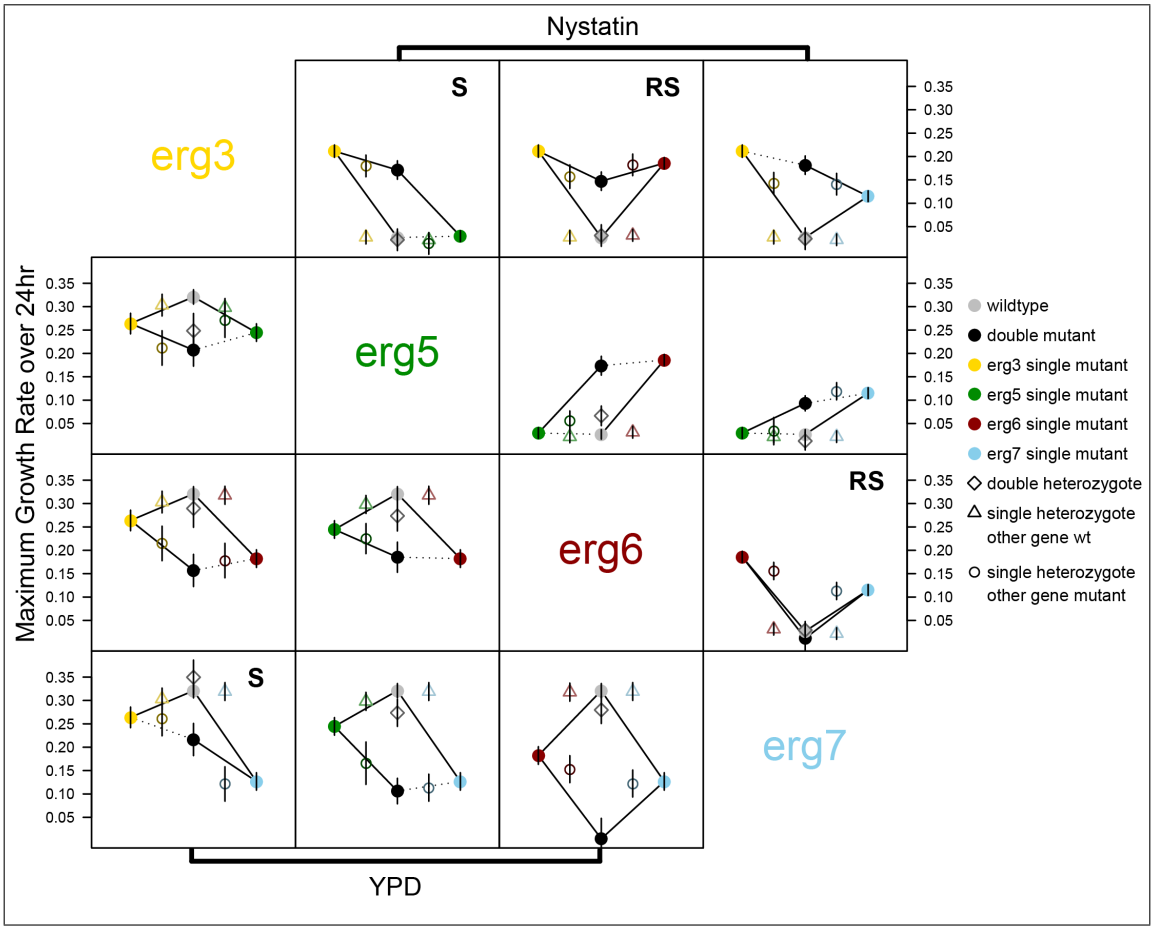

Figure 3.4: Maximum growth rate of diploid strains in nystatin2 (above diagonal) and YPD (below diagonal). Points are the fitted least-squares means of the maximum growth rates, with closed circles determined in the mixed-effects model including only homozygous strains and open symbols from the model that includes heterozygous strains (open diamonds: double heterozygotes; open triangles: single heterozygotes that are wildtype at the other gene; open circles: single heterozygotes that are homozygous mutants at the other gene). Points and bars are otherwise as in Fig 3.3. All symbols are coloured intermediately according to genotype and arrayed along the x-axis so as to lie between the two strains that are genotypically most similar to it. Solid lines indicate significant comparisons in tests run including only homozygous strains while dotted lines are non-significant comparisons. See Fig B.4 for statistical comparisons including heterozygous strains and Fig 3.3 for further graphical details. In YPD, the homozygous comparison *erg3 erg5* vs. *erg3* is not significant when outliers are included. Note that the point for *erg5/ERG5 erg6/erg6* was removed because it was later found to have lost heterozygosity at *ERG5*. All underlying raw data and analyses can be found in Ono *et al.* (2016).

Table 3.3: Results from a mixed-effects model run on all genes using the homozygous diploid maximum growth rate data in nystatin2. For statistical and column details, see Table 3.2.

Term	Coefficient	SE	*P*	Epistasis	Sign	Reciprocal sign
erg3	0.18	0.0065				
erg5	0.0028	0.0058				
erg6	0.16	0.0060				
erg7	0.088	0.0057				
erg3*erg5	-0.043	0.012	**0.00037**	negative	✓	
erg3*erg6	-0.22	0.012	$< \mathbf{10^{-15}}$	negative	✓	✓
erg3*erg7	-0.12	0.012	$< \mathbf{10^{-15}}$	negative		
erg5*erg6	-0.015	0.012	0.19			
erg5*erg7	-0.025	0.010	**0.015**	negative		
erg6*erg7	-0.26	0.014	$< \mathbf{10^{-15}}$	negative	✓	✓

3.3.3 Epistasis for growth in YPD

To determine the extent to which epistasis reflected gross fitness defects not specific to nystatin resistance, we repeated the analysis on maximum growth rate in YPD, a rich growth medium. As in nystatin2, mating type (and its associated auxotrophy) had no significant effect (P = 0.98), and results were averaged over mating types.

The single mutations were generally deleterious in YPD (note the negative coefficients for the individual mutations, Table 3.4 and Table 3.5), consistent with previous characterization of these mutations (GERSTEIN *et al.* 2012). The exception is the haploid *erg5* mutant, which is not significantly less fit than the ancestor in a pairwise comparison of maximum growth rates (bottom left panels in Fig 3.3, Fig 3.4). As observed in nystatin2, the double mutant often had lower fitness than the single mutants in YPD, although the strength of epistasis was generally weak (most interactions resemble a parallelogram, Fig 3.3 and Fig 3.4). Significant sign epistasis was only observed in a single diploid case (*erg3 erg7*).

Epistatic interactions in YPD were qualitatively different from those observed in nystatin2 and often differed between haploids and diploids (Table 3.4 and Table 3.5). In contrast to the prevalence of negative epistasis in nystatin2, significant positive epistasis was observed in some cases (the double mutant is more fit than expected under the additive model). The poor growth in YPD of most double mutant strains suggests that the negative relationships observed in nystatin2 may, in part, be due to intrinsic growth defects, perhaps due to the instability of the cell membrane without proper ergosterol synthesis.

3.3.4 Tolerance across a range of nystatin

To see whether the genetic interactions depended on the concentration of drug, growth was measured as optical density (OD) after 24 hours over a range of nystatin concentrations (0, 1, 2, 4, 8, 16, 32, 64, 128, 256 μM). We focused here on OD to assess the range of environments in which the yeast strain could grow, even if slowly, and because of the massive replication required. While OD is thought to reflect the efficiency of cells'

Table 3.4: Results from a mixed-effects model run on all genes using the haploid maximum growth rate data in YPD. For statistical and column details, see Table 3.2. There were no cases of sign epistasis.

Term	Coefficient	SE	*P*	Epistasis
mating type	0.000042	0.0027	0.98	
erg3	-0.029	0.0057		
erg5	-0.0051	0.0049		
erg6	-0.065	0.0049		
erg7	-0.12	0.0050		
erg3*erg5	-0.026	0.0091	**0.0034**	negative
erg3*erg6	0.0030	0.0090	0.74	
erg3*erg7	0.018	0.0091	**0.041**[a]	positive
erg5*erg6	0.0018	0.0077	0.81	
erg5*erg7	-0.0065[b]	0.0079	0.41	
erg6*erg7	0.040	0.0084	$\mathbf{1.77 \times 10^{-6}}$	positive

[a]Not significant when outliers are included.
[b]Positive when outliers are included.

Table 3.5: Results from a mixed-effects model run on all genes using the homozygous diploid maximum growth rate data in YPD. For statistical and column details, see Table 3.2. There were no cases of reciprocal sign epistasis.

Term	Coefficient	SE	*P*	Epistasis	Sign
erg3	-0.057	0.011			
erg5	-0.076	0.010			
erg6	-0.14	0.010			
erg7	-0.19	0.010			
erg3*erg5	0.020	0.022	0.36		
erg3*erg6	0.032	0.022	0.14		
erg3*erg7	0.15	0.022	$\mathbf{8.5 \times 10^{-11}}$	positive	✓
erg5*erg6	0.079	0.020	$\mathbf{5.6 \times 10^{-5}}$	positive	
erg5*erg7	0.056	0.018	**0.0021**	positive	
erg6*erg7	0.016	0.025	0.53		

ability to turn nutrients into cellular material rather than the rate of growth, OD and maximum growth rate were correlated for the single mutants analysed here (GERSTEIN *et al.* 2012), and the interactions observed were qualitatively similar for the concentrations of nystatin used in both the maximum growth rate and OD assays (0 μM and 2 μM).

As before, mating type was not found to have a significant effect on OD in the haploid data (linear model that included mating type, concentration of nystatin and strain identity as fixed effects; mating type: F = 0.23, df = 1, $P = 0.63$; concentration of nystatin: F = 600.12, df = 1, $P < 10^{-15}$; strain: F = 31.95, df = 10, $P < 10^{-15}$), and data were pooled across mating types.

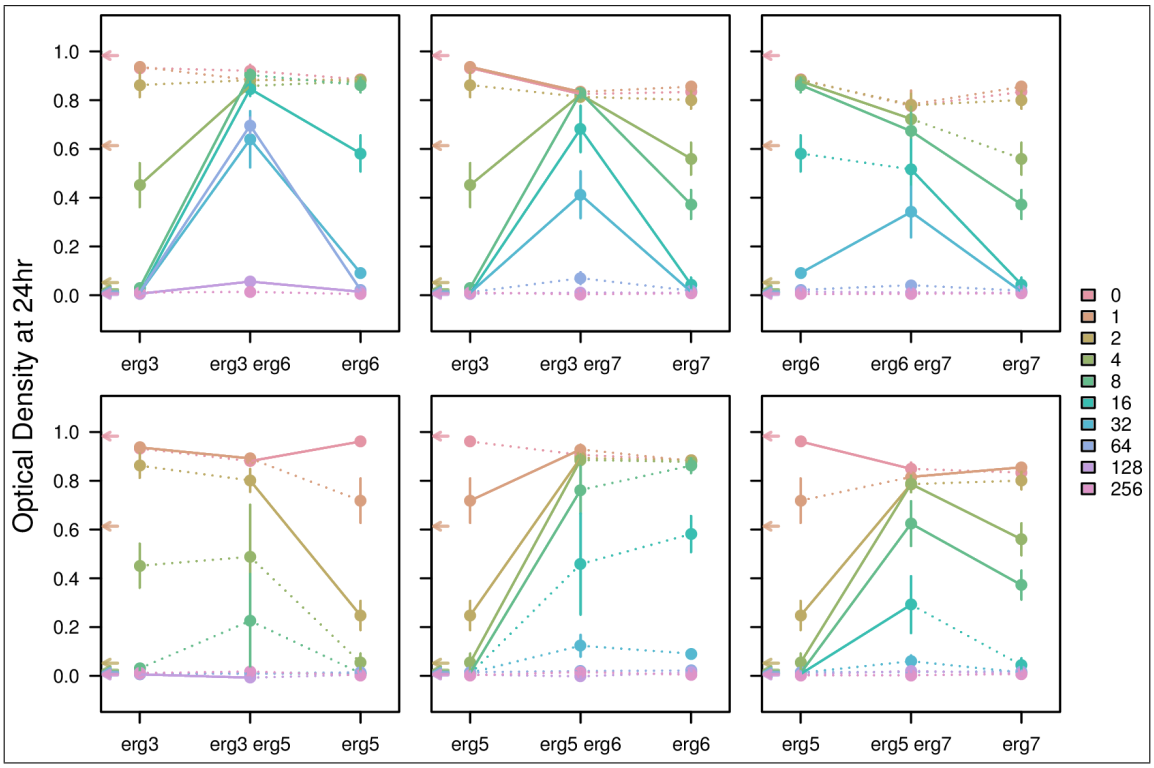

Figure 3.5: Optical density after 24 hours of growth for haploid strains in a range of concentrations of nystatin. Colours go from red to purple, through blues, from lowest to highest concentrations of nystatin. Lines connect different mutants in the same concentration of nystatin. Lines are solid when the difference in OD is significant in a Welch's t-test and dotted when non-significant (not adjusted for multiple comparisons). Arrows on the y-axes indicate the OD of the ancestral strain. Error bars denote the standard error across replicates. All underlying raw data and analyses can be found in ONO *et al.* (2016).

We found that the form of gene interactions changed when measured over a range of concentrations of nystatin (haploid results: Fig 3.5). As observed previously, the double mutant generally had equivalent or lower growth than the two parent mutants at low concentrations of nystatin (0-4 μM), but at high concentrations (32-64 μM), the double mutant strains became the only strains able to grow well. That is, a preponderance of negative epistasis shifted towards a preponderance of positive epistasis as nystatin concentrations rose. This dependence of the sign of epistasis on the concentration of the drug (not only on the presence or absence of the drug) indicates that the outcome of mutation or hybridization will depend heavily

on the specifics of the environment in which the yeast is found.

Homozygous diploid strains showed qualitatively similar patterns of growth to the haploid strains, with the exception of the *erg6/erg6 erg7/erg7* double mutant (Fig B.3). When we compared all diploid strains (including heterozygous strains), interesting patterns emerge (Fig B.5). In many cases, the double heterozygous strain exhibited more growth than either single heterozygous strain (as observed by a 'bump' in the middle of the figure), particularly at higher concentrations of nystatin. This may indicate a net beneficial effect of carrying two heterozygous mutations or may reflect an increased potential for loss of heterozygosity (LOH). LOH, in which a locus that is initially heterozygous for a mutant allele becomes homozygous, would be beneficial in our fitness assay because being homozygous for either mutant allele increases growth in nystatin (compare middle point in Fig B.5 to those second from either end). This may have occurred during the course of the fitness assay, affecting our final measures of fitness. LOH was previously observed for the single heterozygous mutants over a 72-hour time scale (GERSTEIN *et al.* 2014), and being heterozygous for two mutations may increase the chance of LOH for at least one of the two. The unexpected increase in fitness in the double heterozygotes may also be indicative of an epistatic interaction providing some benefit to having two heterozygous mutations within the ergosterol pathway compared to full recessivity (i.e., no benefit) with only a single heterozygous mutation (GERSTEIN *et al.* 2014).

3.3.5 Ergosterol phenotypes and map to fitness

To determine whether epistasis for fitness was consistent with the sterol phenotypes exhibited by the strains, we extracted and measured the sterol profile of all *MATa* strains. In ancestral samples, we see the characteristic four-peaked curve between 240 and 300 nm that is produced by ergosterol and the late sterol intermediate 24(28)dehydroergosterol (ARTHINGTON-SKAGGS *et al.* 1999). Only the latter sterol shows an absorption band at 230 nm, allowing quantification of ergosterol, but we found the peak between 200 and 230 nm to be very sensitive to the standard used (e.g., newly mixed heptane and ethanol vs. heptane layer from extraction performed with no yeast cells and ethanol) and thus limit ourselves to a qualitative description of the results.

All of our single mutants show similar results to those presented by GERSTEIN *et al.* (2012) for these same mutants (Fig 3.6). The two potential loss-of-function mutants (*erg3* and *erg6*) also have similar sterol profiles to knockout mutants of these genes (JENSEN-PERGAKES *et al.* 1998; MUKHOPADHYAY *et al.* 2002). Double mutants show a variety of profiles, as can be seen in Fig 3.6. Notably, most double mutants resemble one of the two parent single mutants, with the exception of the *erg6 erg7* double mutant, which is intermediate between the two single mutants in absorbance over much of the measured range (suggesting a mixture of sterols present). All double mutants that include the mutation in *ERG3* tend to show similar profiles to the *erg3* single mutant. Thus, the sterol profiles were not predicted by gene position in the ergosterol biosynthesis pathway (as *ERG6* is upstream of *ERG3*). Furthermore, the similarity in sterol profiles between double and single mutants did not generally predict the patterns observed for maximum growth rate (with the exception of the *erg5 erg7* haploid and diploid, which behaved like *erg7*, and the *erg3 erg7* diploids, which behaved like *erg3*), indicative of a disconnect between sterol profile and fitness.

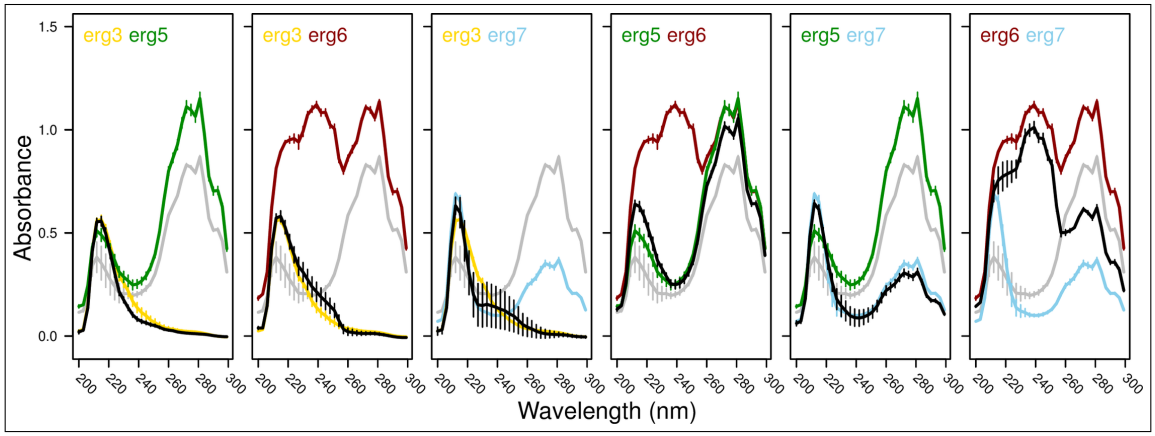

Figure 3.6: Sterol profiles of all *MATa* haploid strains as measured using a spectrophotometry-based assay. The colour scheme is the same as in Fig 3.3, with the double mutant in black and the ancestral strain in grey. Error bars depict the standard error of three replicates with the exception of *erg6 erg7* (2 replicates). The same ancestral and single mutant assays are represented in multiple panels. All underlying raw data and analyses can be found in ONO *et al.* (2016).

3.4 Discussion

We investigated the types of genetic interactions present between pairs of first-step beneficial mutations that arose independently in the presence of the fungicide nystatin. We focused on four mutations, representing each gene found to carry a beneficial mutation among 35 strains evolved in 4 μM nystatin (GERSTEIN *et al.* 2012). All of these genes are in the biosynthesis pathway leading to the production of ergosterol (the primary sterol in the yeast cell membrane, Fig 3.2). When ergosterol is bound by nystatin, the cell membrane becomes permeable to ions, sugars, and metabolites (CARRILLO-MUÑOZ *et al.* 2006), and cell death results. When assayed at 2 μM nystatin, the interactions found among these beneficial mutations were predominantly negative, with double mutants exhibiting a lower growth rate in nystatin than expected based on the combined benefits of the single mutations. This negative epistasis was observed in both haploids (Fig 3.3) and homozygous diploids (Fig 3.4), supporting previous findings that interactions between mutations in functionally related genes are often negative (COSTANZO *et al.* 2016).

3.4.1 Prevalence of sign epistasis

We find that the interactions were so negative that the double mutant grew less well than at least one of the parent single mutants (sign epistasis) in four of the six gene combinations assayed in haploids. In half of these cases, the double mutant grew significantly less well than both single mutants (reciprocal sign epistasis). Similar interactions were observed in diploids (three cases of sign epistasis, two of which were reciprocal). The observation of reciprocal sign epistasis is of particular interest, as this type of BDM incompatibility underlies postzygotic reproductive isolation among speciating lineages. The high frequency of reciprocal sign epistasis observed, even among first-step beneficial mutations acquired in the same environment, confirms the possibility that isolated populations experiencing similar selective pressures can diverge

and eventually speciate simply through the order of mutations that happen to arise and fix (mutation-order speciation).

3.4.2 Maximum growth rate in one environment does not predict sterol phenotype or growth in other environments

The prevalence of sign epistasis among our specific set of beneficial mutations is somewhat surprising given the linearity of the biosynthetic pathway in which all of the affected ergosterol genes act (Fig 3.2). Our results were not consistent with the expectation that the phenotype and fitness of double mutants would be determined by the upstream mutation. In terms of phenotype, the sterol profile of the double mutant was similar to that of the most upstream mutant in only two cases (the *erg5 erg7* and *erg3 erg5* double mutants, Fig 3.6). In terms of fitness, the growth rate of the double mutant differed significantly from that of the most upstream single mutant in three (haploids) and four (diploids) out of six pairwise comparisons (Figs 3.3 and 3.4).

In the combination of two loss-of-function type mutations (*erg3 erg6*), neither sterol phenotype nor fitness matches that of the upstream mutation. These results indicate that there remain substantial interactions between the mutations in the ergosterol pathway, potentially due to partial activity of the upstream genes creating low levels of substrate for the remainder of the pathway, due to downstream genes acting on alternative sterol substrates, or due to interactions among the intermediate sterols themselves. From previous work in the yeasts *S. cerevisiae* and *Candida albicans*, it has been shown that *ERG6* plays a role in offshoot sterol synthesis in mutants of *ERG3* (SANGLARD *et al.* 2003), and it is known that intermediate sterols are found at different levels in different compartments of the cell (ZINSER *et al.* 1993) and may impact fitness in a variety of ways (e.g., altering temperature tolerance CASPETA *et al.* 2014 and virulence MCCOURT *et al.* 2016).

There was also no clear relationship between sterol phenotype and fitness in these strains. Sterol phenotype for most double mutants resembles one of the two single mutants (Fig 3.6), but this similarity in sterol phenotype did not generally predict maximum growth rate in nystatin2 (with the possible exception of *erg5 erg7* in haploids and diploids and *erg3 erg7* in diploids, Figs 3.3 and 3.4). Future analyses that determine the processing of sterols in the single and double mutants, as well as their pleiotropic effects, would further elucidate these genetic interactions.

Interestingly, the type of epistasis depended strongly on the concentration of nystatin. At lower concentrations of nystatin, similar to those used to acquire the mutations ($\leq$4 μM nystatin), epistatic interactions were typically negative (Fig 3.5), with the double mutant showing similar or lower densities after 24 hours of growth than the single mutants. By contrast, at higher concentrations of nystatin, the interactions were often positive, with double mutants typically able to outgrow both single mutants. Emblematic of this phenomenon, the best growing haploid double mutant strains at 32 μM nystatin (*erg3 erg6, erg3 erg7, erg6 erg7*) were also those that exhibited the most negative epistasis at lower concentrations. This implies a tradeoff between growth in a lower concentration of the fungicide and tolerance to high concentrations of the drug. Conceptually, this tradeoff suggests that the double mutant initially overshoots the optimum when nystatin concentrations are low, because the costs associated with each ergosterol mutation are combined (perhaps destabilizing the plasma membrane); by contrast, when nystatin concentrations are high, the optimum is

shifted even further away, and extreme reductions in ergosterol and potentially other sterols are needed for the yeast to survive, at which point the double mutant is most fit (see, e.g., BLANQUART *et al.* 2014 for a theoretical exploration of this phenomenon). Because membrane damage can trigger cell cycle arrest in yeast (KONO *et al.* 2016), another possible explanation for the results observed at high concentrations of nystatin is that single mutants experience cell cycle arrest, reducing growth rate, whereas the additional stress caused by the combination of two mutations and high concentrations of nystatin may cause a checkpoint failure in double mutants, allowing the cells to bypass arrest and continue dividing (C. Nislow, pers. comm.).

The shifting nature of epistasis as a function of the severity of the environment also has implications for speciation and has not been widely discussed (but see KISHONY and LEIBLER 2003 for discussion about environment-dependent epistasis and ARNEGARD *et al.* 2014 for an example of an environment-dependent negative epistatic interaction on feeding and growth performance in F2 hybrid stickleback). Our results show that BDMs can be environment-specific, and thus gene flow between species might vary according to the environment in which secondary contact occurs (BORDENSTEIN and DRAPEAU 2001). Counterintuitively, our results further suggest that harsher environments may be more conducive to gene flow because of the possible benefit of combining adaptive mutations from different populations. Indeed, environments that are so harsh that only strains combining mutations can survive (as we observed at high concentrations of nystatin) might promote hybridization and potentially lead to hybrid speciation (reviewed in MALLET 2007). For example, extreme desert environments have selected for combinations of traits that improve drought tolerance, allowing hybrid *Helianthus* sunflowers to colonize and proliferate (GROSS and RIESEBERG 2005).

3.4.3 Fitness landscapes in haploids and diploids

Because cell volume to surface area ratios are different for haploids and diploids (MABLE 2001), we might expect differences in growth and epistasis between haploids and diploids, particularly in the face of a selective pressure like nystatin that impacts the cell membrane. By and large, however, our results were consistent regardless of ploidy, with diploid homozygous mutants and haploid mutants showing similar patterns of epistasis. One exception was the *erg6/erg6 erg7/erg7* double mutant, which was so unfit in diploids that yields were often too low to obtain initial cell densities similar to other strains in our growth rate assays, even after extended incubation. The haploid version of this same double mutant, however, also showed very low fitness. During the initial isolation of the haploids from spores, the double mutant colonies were identifiable by their noticeably smaller size compared to those produced by single mutant and ancestral genotypes. A haploid double mutant strain also exhibited reversion in one instance during growth in 10 mL YPD (Sanger sequencing revealed a secondary mutation in the same codon as the original mutation, reverting the amino acid).

Considering the various diploid heterozygotes, we confirmed that the ergosterol mutations were largely recessive, as found previously for the single heterozygous mutant strains (GERSTEIN *et al.* 2014). There were more signs of nystatin resistance in the double mutant strains than in the single mutants, however. One indication of this was the double heterozygous strains showing a slight increase in biomass produced (as measured by optical density) compared to the single heterozygous strains across a range of concentrations of nystatin (Fig B.5). Despite this, the double heterozygous strains were uniformly of low growth rate in

nystatin2 (open diamonds in Fig 3.4), with similar sensitivity as found in the ancestor. The generally poor performance of the double heterozygous diploid is of particular interest because this genotype would be the first hybrid product of crosses between strains fixed for different beneficial mutations (see also Fig B.4). Thus, F_1 hybrid inviability in the double heterozygotes, as well as reciprocal sign epistasis, contributes to reproductive isolation between these strains.

3.4.4 Implications for speciation

Overall, we find that the very earliest stages of divergence within a common selective environment can generate postzygotic reproductive isolation, observing sign epistasis, reciprocal sign epistasis, and F_1 hybrid inviability in double heterozygotes among the first-step adaptive mutations isolated in the presence of nystatin. Although we did not assay incompatibilities at other stages (e.g., meiotic incompatibilities), we expect that further BDMs might be revealed by analyzing other stages in the life cycle (indeed, it was very difficult to sporulate some double mutant strains, particularly *erg5/ERG5 erg7/ERG7*). We speculate that genetic incompatibilities may be especially likely in scenarios such as the one investigated here, where selection favours large effect mutations. In the initial experiment in which mutations were acquired, the concentration of nystatin was chosen to inhibit growth, so that only mutations capable of rescuing fitness were isolated (GERSTEIN *et al.* 2012). Such large effect mutations might have more costly pleiotropic effects and/or be more likely to overshoot the fitness optimum when combined, showing negative epistasis for fitness even if their effects are multiplicative or additive on the underlying trait. If large effect mutations are more likely to interact negatively, which is consistent with our results and others (CHOU *et al.* 2011; KHAN *et al.* 2011; KRYAZHIMSKIY *et al.* 2014; SCHENK *et al.* 2013), short periods of severe selection might be more likely to lead to speciation than longer periods of mild selection. Future experiments comparing genetic incompatibilities among strains with similar levels of divergence but consisting of a few large effect or several small effect genetic differences would be extremely valuable. We also speculate that independent populations experiencing directional selection to the same environmental change might be more likely to speciate than those experiencing directional selection to different environments because the beneficial mutations that accumulate in the former case may be more likely to involve similar pathways and thus more likely to interact negatively (as has been shown in interaction studies, see COSTANZO *et al.* 2016; HARTMAN *et al.* 2001). Indeed, even though the beneficial mutations that we assayed were all in the same 'linear' pathway and acquired in the same selective environment, we found that the type of epistasis that underlies speciation was common, providing experimental support for the mutation-order speciation hypothesis.

Chapter 4

The limit to evolutionary rescue depends on ploidy in yeast exposed to nystatin

4.1 Introduction

As their environment changes, species must adapt to persist. By knowing the factors that affect population size and adaptation, we can begin to predict or influence a population's likelihood of survival. Evolutionary rescue occurs when a population is saved from eventual extinction in response to an environmental change by genetic adaptation to this new environment (CARLSON *et al.* 2014). Broadly, evolutionary rescue is important to understand from two main standpoints. The first is conservation - where the question is whether a species will be able to survive current and future environmental changes. The second is related to human health - where the question is whether pests and pathogens are able to evolve to overcome human intervention. In evolutionary rescue, populations increase their growth rate either using existing standing genetic variation, new alleles introduced by gene flow, or new mutation. Populations reach their adaptive limit when they are not able to acquire the genetic changes necessary for rescue before going extinct. In this paper, we focus on the case of evolutionary rescue from new mutation, asking: Does ploidy influence the limits to adaptation?

The genomic characteristics of the organism may influence the potential for evolutionary rescue in a population. For example, epistasis between potential adaptive mutations and the genetic background can affect which mutations are actually beneficial. Here, we focus on ploidy as a major genomic factor that can impact the outcome of evolution. For a variety of reasons, diploid and higher ploidy populations are expected to have different rates of evolutionary rescue from haploid populations (OTTO and WHITTON 2000). With diploidy, the dominance of adaptive mutations can strongly influence the likelihood of rescue because recessive mutations will not be 'seen' by selection when they initially arise in heterozygous form. Recessive mutations will then have a smaller chance of rising to high frequency in the population and will take longer to do so compared to partially dominant or fully dominant mutations (ORR and OTTO 1994). However, many pests and pathogens are either haploid (for example, infectious bacteria), alternate between haploid and diploid phases or have haploid and diploid individuals (for example, the spider mite pest *Tetranychus urticae* has haploid males and diploid females). For these organisms, recessive mutations can be immediately beneficial if they have a positive selection coefficient in the haploid individuals/phase and can aid in the evolution of antibiotic or pesticide resistance. On the other hand, diploids should acquire twice the number of mutations, if haploids and diploids have the same per basepair mutation rate, which may allow diploids to be rescued when mutations are sufficiently dominant (e.g., OTTO and WHITTON 2000), especially when adaptation is

limited by mutational availability (ORR and OTTO 1994). Experiments performed using yeast support this prediction. In large populations, adaptation is faster in haploids because mutations are not limiting, but when population sizes are decreased, the haploid advantage is lost (ZEYL *et al.* 2003).

The type of environmental change also determines the ease of evolutionary rescue from new mutation. If a higher fraction of possible mutations are adaptive in the new environment, evolutionary rescue should occur more frequently because any individual mutation will have a higher chance of being beneficial. A related factor may be the rate of environmental change. In experimental populations of *Escherichia coli* adapting to the antibiotic rifampicin, LINDSEY *et al.* (2013) determined that rapidly changing environments not only limit the number of available mutations by lowering population size but also make certain evolutionary trajectories inaccessible because multiple mutations are required, and these mutations are not all individually beneficial in high concentrations of the drug. As a result, there is a much smaller fraction of surviving populations when there are faster rates of environmental change.

Little is known about the effect of ploidy on evolutionary rescue from new mutations and the topic has not been well-discussed in the theoretical literature (reviewed in ALEXANDER *et al.* 2014), thus making it an interesting avenue of exploration. Most investigations have focussed instead on factors such as the rate of population size decline with environmental stress and the relative contribution of new mutations vs. standing genetic variation. Some models of evolutionary rescue are designed with or can be extended to diploid populations but they tend to assume that the dominance of rescuing mutations is relatively high, giving the heterozygote a fitness advantage over the ancestral type (e.g., GOMULKIEWICZ and HOLT 1995, ORR and UNCKLESS 2008). When evolutionary rescue is allowed to occur from standing genetic variation, as opposed to new mutations, dominance is not as important because the necessary mutation may already be present at a high enough frequency to make homozygotes common (ORR and UNCKLESS 2008). In fact, if the rescue mutation was deleterious prior to the change in environment and populations are started from mutation-selection balance, a recessive allele is as likely to fix as a dominant allele because more copies of the recessive allele will already be present in the population (ORR and BETANCOURT 2001). If standing genetic variation is produced through hybridization between the focal population and either a divergent population or a closely related species, diploidy may even be favoured if beneficial mutations are often dominant or overdominant (as observed between species of yeast, BERNARDES *et al.* 2017) and deleterious mutations or incompatible alleles between groups are recessive (predicted to be the case between yeast species based on STELKENS *et al.* 2014). Less is known about the dominance of beneficial mutations in general, however, which is important for populations of a diploid organism relying on new mutations for evolutionary rescue. Much of the relevant experimental work has been done with bacteria (e.g., studies of antibiotic resistance evolution reviewed in MACLEAN *et al.* 2010) or haploid eukaryotes (e.g., haploid yeast in BELL and GONZALEZ 2009 or primarily haploid *Chlamydomonas reinhardtii* in LACHAPELLE and BELL 2012), thus a study of the effects of ploidy on evolutionary rescue will bring novel insights.

We investigate the ability of yeast, *Saccharomyces cerevisiae*, to undergo evolutionary rescue in response to a high concentration of the fungicide nystatin and how the ploidy of the yeast affects this ability. We model our design after a previous experiment in this system performed exclusively with haploids (GERSTEIN *et al.* 2012), which found that the adaptive alleles acquired by haploid yeast were all recessive in the diploid state

(GERSTEIN *et al.* 2014). These recessive alleles will not be sufficient to rescue diploid populations of yeast, as they will appear in heterozygotes, but the mutations found by GERSTEIN *et al.* (2012) are unlikely to be the only possible beneficial mutations in this environment, potentially missing dominant beneficial alleles. Diploids, aided by their larger genome size and therefore higher number of mutations per cell, may explore a wider range of potentially adaptive mutations, finding some that are at least partially dominant. For these reasons, we expected diploid yeast to access alternative evolutionary paths to the haploids, albeit diploids would likely undergo evolutionary rescue less frequently due to the apparently low availability of dominant beneficial mutations in this environment. Supporting this idea is a similar study performed by ANDERSON *et al.* (2004) that evolved both haploid and diploid yeast to the drug fluconazole, for which candidate pathways to drug resistance are known. At low concentrations of the drug, mutations in *PDR1* and *PDR3* were favoured, and diploids were able to adapt faster than haploids due to increased mutation availability and therefore decreased waiting time. All mutations found in diploids were dominant (29/29) while only about half of those found in haploids were dominant when tested in a heterozygous diploid background (12/29) despite almost all being found in one of those two genes. At high concentrations of the drug, however, recessive mutations in *ERG3* (one of the genes also implicated in nystatin resistance) are favoured, and diploids were found to evolve slower than haploids (ANDERSON *et al.* 2004). The authors concluded that diploids likely required two mutational events to occur (a mutation in *ERG3* followed by a second mutation that rendered the first mutation in *ERG3* homozygous) in order to acquire resistance, slowing their adaptation. In the current study, we set out to determine the mutations involved in the evolutionary rescue of diploid populations to a high concentration of nystatin and found that, by and large, diploid populations did not genetically adapt within our short time course evolution experiments, even though we observed hundreds of cases of rescue in haploid populations over this same time period. These results have implications for the efficacy of nystatin when applied to fungal pathogens, which include *Candida albicans*, a common human fungal pathogen that is predominantly diploid (HICKMAN *et al.* 2013), and *S. cerevisiae*, for which clinical strains are mostly either diploid or of higher ploidy (ZHU *et al.* 2016) .

4.2 Methods

4.2.1 Strains

In total, three mutant acquisition experiments, similar to GERSTEIN *et al.* (2012), were performed, plus one flask experiment conducted at large population size. Except where noted, we used the S288C background, using the strains BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*), BY4739 (*MATα leu2Δ0 lys2Δ0 ura3Δ0*) (Open Biosystems) and a diploid produced by mating the two (BY4741xBY4739, mating described in ONO *et al.* 2017). To assess the sensitivity of the results to strain background, Acquisition Experiment 3 was performed using a different background (W303), using the haploid strains MJM64 (*MATa-YCR043C::KANMX STE5pr-URA3 ade2-1 his3Δ::3xHA leu2Δ::3xHA trp1-1 can1::STE2pr-HIS3 STE3pr-LEU2*) and MJM36 (*MATα-YCR043C::HPHB STE5pr-URA3 ade2-1 his3Δ::3xHA leu2Δ::3xHA trp1-1 can1::STE2pr-HIS3 STE3pr-LEU2*) constructed by MCDONALD *et al.* (2016), and a diploid produced by mating (OLY075). The diploid was generated following the procedures of MCDONALD *et al.* (2016) with

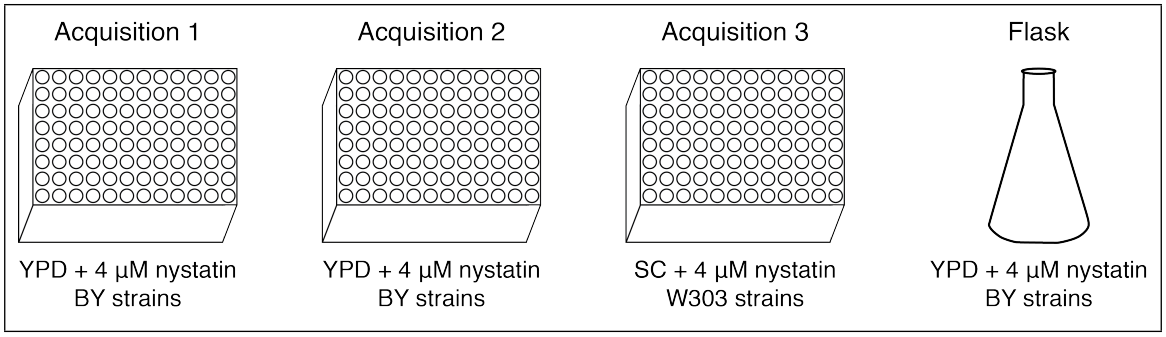

Figure 4.1: Visual representation of all mutant acquisition experiments. The first two experiments were performed in 96-well deep well boxes using the BY strains (BY4741, BY4739 and BY4741xBY4739) and YPDnystatin4 as the medium. For two wells from Acquisition Experiment 1 and three wells from Acquisition Experiment 2, the protocol was not followed correctly, and these wells are excluded from the paper entirely. The third acquisition experiment was performed similarly except that it used W303 strains (MJM64, MJM36 and OLY075) and SCnystatin4. In this experiment, due to space constraints, 80 wells were inoculated from the same pre-growth culture as another well in the experiment (all *MAT*α). In the analysis, we restrict ourselves to considering only one of these two replicates because they are not independent. If neither replicate grew, we counted that as no growth (63 cases). If one of the two replicates grew, we counted that as growth and performed further assays on the population that grew (13 cases). If both replicates grew, we chose one at random to analyze and discarded the other (4 cases). The final acquisition experiment investigated a much larger population size (roughly 100-fold greater), being performed in flasks instead of deep well boxes and again using the BY strains and YPDnystatin4.

slight modifications; using YPAD in place of YPD (YPD + 40 mg/L adenine sulfate), incubating the matings overnight, excluding the PBS buffer step, and performing the selection step twice. The growth of all strains was inhibited under the treatment conditions, with no observable positive growth in the absence of a resistant mutant within 48 hours (Fig. C.2).

4.2.2 Mutant acquisition in deep well boxes

For an overview of all acquisition experiments performed, see Fig. 4.1. For Acquistion Experiments 1 and 2, the strains were first struck from frozen on YPD plates and grown at 30°C. After three days of growth, 100-well honeycomb plates used with the Bioscreen C Microbiological Workstation (Thermo Labsystems) were filled with 150 μL of YPD per well, and each well was inoculated with a separate colony to produce stationary phase cultures for use in the mutation acquisition phase. The plates were grown for 24 hours in the Bioscreen machine at 30°C with maximum, continuous shaking. Note that all liquid medium throughout this paper was supplemented with ampicillin at a final concentration of 0.04 mg/mL to prevent bacterial growth. In Acquisition Experiment 1, we used 191 colonies of the diploid strain and 191 colonies of the haploid strains, split between mating types (*MATa*: 96, *MAT*α: 95), alternating between diploid and haploid strains throughout the plates. In Acquisition Experiment 2, we used 286 colonies of the diploid strain, 47 colonies of *MATa*, and 48 colonies of *MAT*α, with any given plate containing either all diploids (three plates) or half *MATa* and half *MAT*α (one plate).

The following day, the honeycomb plates were visually assessed to confirm full growth in each well within YPD. 10 μL from each well was used to inoculate 990 μL of YPD + 4 μM nystatin ('YPDnystatin4') in deep well boxes, after mixing the culture by pipetting up and down. We estimate this inoculum to contain ~7.0×10^5 cells per well (based on hemacytometer counts). The same general map was used for the boxes as for the honeycomb plates. The boxes were covered with aluminum lids to prevent cross-contamination while sampling and plastic lids were added on top to protect the aluminum lids. They were incubated at 30°C, shaking at 200 rpm.

The wells were checked every 24 hours for growth by visual examination. Clumps were typically observed at the bottom of the well prior to the yeast covering a larger, circular area, so a well was considered to have growth if the bottom was covered with enough yeast cells to form two small clumps, about 2 mm long and 1 mm wide. On the first day that growth was observed in a well, the growth was recorded, the aluminum lid was sterilized with 70% ethanol and punctured with a pipet tip, and the culture was frozen in 15% glycerol at -80°C. In Acquisition Experiment 1, 42 *MATa* strains, 86 *MAT*α strains and 90 diploid strains were collected over the course of 12 days. In Acquisition Experiment 2, 16 *MATa* strains, 48 *MAT*α strains and 100 diploid strains were collected over the course of 10 days.

Acquisition Experiment 3, performed in the W303 background, used similar methods to Acquisition Experiment 1, with exceptions described here. The strains were originally struck from frozen on YPAD plates and grown for only two days. SC (supplemented with adenine) was used instead of YPD for growth in the honeycomb plates. SC was formulated using 20 g/L of dextrose, drop-out mix complete (US Biological, D9515), and yeast nitrogen base including ammonium sulfate, according to the manufacturer's instructions. This medium was supplemented with an additional 57 mg/L of adenine sulfate. There were 155 colonies of the diploid strain, 80 colonies of the *MATa* strain and 165 colonies of the *MAT*α strain used for the growth in the honeycomb plates, alternating between strains throughout. SC + 4 μM nystatin ('SCnystatin4') was used in place of YPDnystatin4 in the deep well boxes. Initial inoculum was estimated as 7.4×10^5 cells per well. Pilot experiments indicated that mutants would be difficult to isolate in the *MAT*α background (the initial stock was later found to be respiratory-deficient), so 80 additional wells of this strain were added to the deep well boxes, using the same pre-growth culture as for one other well. A total of 245 wells of the *MAT*α strain were included in the experiment, so the map of the deep well boxes was slightly modified from that of the honeycomb plates, although still alternating between strains. Over the course of seven days, 77 *MATa* strains, 34 *MAT*α strains and 121 diploid strains were collected.

4.2.3 Confirming nystatin resistance

All populations frozen from the acquisition experiments were tested for resistance to nystatin. Populations were pre-grown from frozen by inoculating 975 μL of 0.5 μM nystatin (in YPD for Acquisition Experiments 1 and 2, SC for Acquisition Experiment 3) with 25 μL of frozen culture in deep well boxes according to a randomized map. Aluminum lids were added to the boxes, and they were incubated at 30°C, shaking at 200 rpm for 72 hours, after which almost all wells had full growth as judged by visual inspection. From these pre-growth cultures, 200 μL was transferred to a 1.5 mL tube and stored at 4°C.

We measured growth in YPDnystatin4 (for Acquisition Experiments 1 and 2) or SCnystatin4 (for Acqui-

sition Experiment 3) of all populations on three different days using the Bioscreen C, which automatically measures optical density ('OD') in 100-well honeycomb plates. For each growth assay, honeycomb plates filled with 148.6 μL of nystatin medium were inoculated with the cultures kept at 4°C according to a new random map for each day. The culture tubes were vortexed until fully resuspended, and 1.5 μL was transferred to the appropriate well. All tubes were returned to 4°C when inoculation was completed. OD was measured automatically using the wideband filter at 30-min intervals for 72 hours from cultures growing at 30°C with medium continuous shaking.

Lines were considered potentially mutant when the well had an OD after 72 hours of growth ('OD72') that was greater than the halfway point between the well with the lowest OD and that with the highest OD, including the ancestral controls, in the majority of assays (at least two out of three). We consider these to be 'potentially' mutant because growth of the yeast can depend on the exact conditions of the assay (including plate type and volume of medium), thus they are not necessarily fully resistant to the original evolutionary conditions (deep well box, 1 mL of medium). Based on initial testing, we knew that our collection of haploid populations included some nystatin-resistant mutants, so the highest OD reflected a truly resistant strain. A single cutoff was used for both Acquisition Experiment 1 and 2 for each assay day since they were tested in the same medium and always assayed together. Because Acquisition Experiment 3 used a different medium and the assays were conducted on different days, the cutoffs were recalculated.

4.2.4 Mutant acquisition with larger population sizes

Because we had such difficulty finding diploid mutant strains using relatively small population sizes (roughly 7×10^5 cells inoculated) in our original acquisition experiments, we conducted a mutant acquisition experiment with a higher population size (Fig. 4.1). To do so, we first struck BY4741, BY4739 and BY4741xBY4739 from frozen onto YPD plates and allowed them to grow for three days at 30°C. We then inoculated 10 mL of YPD in separate test tubes for colonies from 10 diploid colonies and one of each haploid mating type. These tubes were allowed to grow for 24 hours in a rotor at 30°C. The next day, we inoculated 250 mL flasks filled with 99 mL of YPDnystatin4 with 1 mL of the overnight culture (12 flasks in total). We estimate this inoculum to contain ~7.0×10^7 cells (based on counts in a hemacytometer).

The flasks were covered with aluminum foil and incubated at 30°C, shaking at 200 rpm. The flasks were checked every 24 hours by visual examination for growth. When growth was observed (as a noticeable lightening of the culture colour and loss of clarity when compared to a flask containing no yeast), 500 μL of culture was sampled and frozen at -80°C in 15% glycerol. All flasks showed growth by Day 10. Resistance to nystatin was then assayed as described for the first two acquisition experiments with cutoffs being recalculated because these assays were performed on different days.

4.2.5 Further testing of potential diploid mutants

All 13 diploid populations that consistently showed growth in nystatin and underwent further testing had the BY genetic background since all were from Acquisition Experiments 1 or 2. Because initial results indicated that some 'diploid' mutants were actually haploid contaminants, all potential diploid mutants were verified by replica plating. Haploids of this background carry auxotrophies and can thus be detected both

by their ability to mate with haploids of the opposite mating type (tested by mating with haploid strains carrying different auxotrophic mutations and testing for complementation) and by their inability to grow on either SC lacking histidine (*MATa*) or SC lacking lysine (*MAT*α). In this way, we determined that nine of the populations were primarily composed of haploid contaminants, and these were excluded from further analyses. Four diploid potential mutant populations remained after these tests.

These four diploids, along with ancestor and haploid mutant controls, were further tested for growth in YPDnystatin4. Yeast were struck from frozen on YPD plates and incubated at 30°C for three days. Between six and eight colonies per population were chosen haphazardly, picked into 1 mL of YPD in deep well boxes, and incubated at 30°C, shaking at 200 rpm overnight. From this culture, 10 μL was used to inoculate 990 μL of YPD and another 10 μL was used to inoculate 990 μL of YPDnystatin4, both in deep well boxes. Aluminum lids were added to these boxes to minimize evaporation. The boxes were incubated at 30°C, shaking at 200 rpm for 72 hours and visually inspected for growth every 24 hours. Growth was scored on a scale from zero (no growth) to three, where one is barely recognizable growth (a light dusting of cells or a circle of roughly 1 mm diameter), two is recognizable but not full growth (a patchy covering of cells or a circle up to roughly 3 mm diameter) and three is full growth (a circle of greater than roughly 3 mm diameter up to a full covering of the bottom of the well). Wells were chosen for freezing based on their growth after the 72 hour incubation. For the populations that showed full growth in YPDnystatin4, one well was chosen randomly for freezing (coming from a single colony). If a population showed variable growth across colonies, the colony showing the best growth was chosen for freezing, or one of the colonies showing the best growth was chosen randomly. Colonies were frozen from the first YPD box in 15% glycerol at -80°C.

A growth assay was performed using the Bioscreen C machine on the colonies chosen for freezing, along with randomly chosen colonies from the strains that did not show growth in YPDnystatin4, using culture from the second YPD box (inoculated on the same day as the YPDnystatin4 box). 1.5 μL of culture was used to inoculate 148.6 μL of both YPD and YPDnystatin4 in four wells each, using freshly diluted nystatin stocks. OD was measured automatically using the wideband filter at 30-min intervals for 72 hours from cultures growing at 30°C with medium continuous shaking. No diploid colony showed growth in this growth assay, and it was later determined that these four diploid populations also harboured haploid contaminants at very low frequencies. This was determined by testing the culture from the end of the original growth assays as described at the beginning of this section.

4.2.6 Nystatin efficacy over time

While many diploid cultures appeared to undergo evolutionary rescue in nystatin, growth was typically only observed late in the acquisition experiment, and most of these populations did not exhibit resistance when regrown in nystatin. We thus tested the ability of nystatin to inhibit yeast growth over a time course to investigate whether the effective concentration of the drug was decreasing over the duration of the acquisition experiments. We also tested whether the presence of dead yeast cells altered the efficacy of nystatin. Dead cells were obtained by growing cultures of BY4741xBY4739 in 1 mL of YPD in 1.5 mL tubes overnight at 30°C, shaking at 200 rpm. The next day, these tubes were placed in a heat block at 95°C for 10 minutes to

kill the cells. Preliminary tests indicated that this amount of time was sufficient to kill all of the cells. On the first day of the experiment (Day 0), 990 μL of YPD, YPDnystatin4, SC and SCnystatin4 were placed in deep well boxes, with half of all nystatin-containing wells also receiving 10 μL of heat treated cells, giving a total of 16 wells per type (base medium, base medium with nystatin, and base medium with nystatin and dead cells) per box. Ten μL of live BY4741xBY4739, grown overnight in 1 mL of YPD, was used to inoculate 120 wells on each of Day 0, 4, and 8 of the experiment, split evenly across boxes and medium types. On Day 12, there was not enough live culture to inoculate all wells. Many control (YPD or SC) wells were not inoculated. One SCnystatin4 + dead cells well was not inoculated and excluded from the analysis. The other wells were inoculated with as much as possible, ranging from 2 μL to 10 μL of culture (the fact that fewer cells were used on Day 12 is conservative with respect to our results below). Between inoculation days, the stocks were stored at 4°C while the deep well boxes were incubated at 30°C, shaking at 200 rpm. The boxes were visually inspected for growth every 24 hours for 16 days, excluding Days 10 and 11, using the same scoring system from zero to three as described in Section 4.2.5.

4.3 Results

Including all four acquisition experiments, we found that none of the 619 inoculated diploid wells (excluding those found to have haploid contamination) or flasks underwent evolutionary rescue, compared with 116 out of 223 *MATa* wells (52%) and 100 out of 308 *MAT*α wells (32%) (Fig 4.2). These numbers are based on tests of resistance to nystatin performed by following the growth of the populations for 72 hours in their original medium type (either YPDnystatin4 or SCnystatin4) in the Bioscreen C. We conclude that ploidy played a substantial role in the likelihood of evolutionary rescue. For each of the three acquisition experiments in deep well boxes, strain type influenced the proportion of tested populations that were determined to be putatively resistant (χ^2 contingency test, Acquisition Experiment 1: χ^2 = 53.14, df = 2, p-value = 3×10^{-12}; Acquisition Experiment 2: χ^2 = 119.36, df = 2, p-value < 10^{-15}; Acquisition Experiment 3: χ^2 = 193.98, df = 2, p-value < 10^{-15}). *MATa* consistently had the highest proportion of putative mutants among the wells that grew in the original acquisition experiments (1: 26/42 = 0.62; 2: 14/16 = 0.88, 3: 76/77 = 0.99), followed by *MAT*α (1: 42/86 = 0.49; 2: 38/48 = 0.79, 3: 20/34 = 0.59). Only four of the 81 diploid populations tested from Acquisition Experiment 1 were deemed to be putative mutants, and none from Acquisition Experiments 2 or 3 (Fig 4.2 and Fig C.2). These four potential mutant populations may be weakly resistant (allowing them to grow in lower concentrations of nystatin), but they were also determined to harbour haploid contaminants at low frequencies, explaining the growth observed in YPDnystatin4. These results indicate that ploidy restricts the ability of yeast populations to undergo evolutionary rescue under these conditions.

The relative absence of adaptive evolution among diploids is unlikely to be due to a small number of replicates tested. We modelled the growth of a population from a single cell established on a plate to the portion of liquid that is used in the inocula for the acquisition experiments and expect 0.82 nucleotide-changing mutations to have occurred in at least one cell per site within the genome, combining all deep well acquisition experiments and the flask experiment (Section C.2). When we multiply this over the average length of an ORF, we expect to have sampled ~888 different non-synonymous or nonsense mutations for

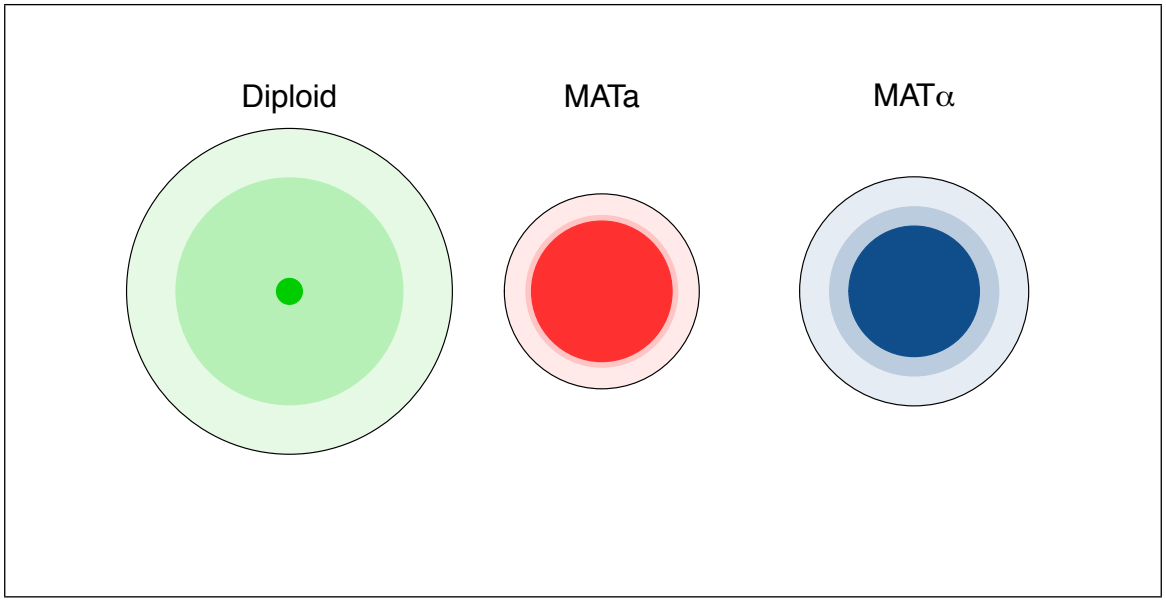

Figure 4.2: Plot indicating the number of potential mutants found in all wells over all three acquisition experiments in the deep well boxes. The area of the lightest coloured circle is proportional to the total number of wells inoculated of each type. The intermediately coloured circle is proportional to the total number of wells that showed some growth in the initial acquisition experiments. The darkest coloured circle is proportional to the number of wells considered to carry potential mutants based on the follow-up growth assays. Very few of the diploid populations grew in the follow-up assays (green; 4/302/623; putative mutants/wells that grew/wells inoculated), while the majority of haploid populations that grew in the acquisition experiments were reliable mutants (red: *MATa* 116/135/223; blue: *MAT*α 100/168/308). Further testing indicated that these four diploid populations harboured haploid contamination at a low frequency.

each gene within the genome over the course of the experiment. Based on these numbers, we believe that if a one-step dominant rescue mutation were available in the diploids, we should have seen full resistance evolve. If a secondary mutation is required in the same gene (either a second mutation in the same ORF or a loss of heterozygosity of the first mutation), there is a <0.0003 chance that such a secondary mutant cell would have been sampled over the course of the experiment (Section C.2). Thus, this experiment tests primarily for single-step rescue mutations and not for the potential effects of homozygous mutations.

4.3.1 Further testing of potential diploid mutants

Four diploid populations that grew in Acquisition Experiment 1 were considered to potentially carry rescue mutations (DR1-4 for "Diploid Rescue"). Two of these (DR1, 2) grew in all three replicate Bioscreen C runs when testing to confirm resistance, while the other two (DR3, 4) grew in only two of the three follow-up assays. When colonies from these populations were retested under the original acquisition conditions (YPDnystatin4 in a deep well box), only a small amount of growth was observed for DR1-3 and no growth was observed for DR4 in 72 hours. One colony from each of these populations (chosen among those that grew for DR1-3) was then assayed for growth in the Bioscreen, this time using freshly diluted nystatin, and no growth was observed for any of these strains in YPDnystatin4 while all had full growth in YPD. The nystatin

used in the previous assays was older and possibly slightly degraded, potentially explaining the observed growth of DR1-4. These results indicate that the populations may harbour weakly resistant or higher fitness mutants, but not mutations that are fully resistant to 4 μM nystatin. Further, follow-up investigation of these populations from the end of the original growth assays indicated that low frequency haploid contamination was present in the original populations.

4.3.2 Nystatin efficacy over time

Based on the observation that many populations grew in the initial acquisition experiments but were not resistant when re-tested in nystatin (see Section C.1), we hypothesized that nystatin was losing efficacy over the course of the acquisition experiments (lasting seven to twelve days). Both YPDnystatin4 and SCnystatin4 showed significant degradation of nystatin efficacy over time (Fig 4.3). This was observed as a loss of the ability to inhibit the growth of BY diploids within four days. The presence or absence of dead yeast cells in the medium did not have a significant effect on this loss of efficacy but the base medium did, with SC allowing more growth overall than YPD (generalized linear model with a binomial error distribution and logit link function performed using the *glm* function in the package *stats* in R [R CORE TEAM 2016], p-values determined based on sequential likelihood ratio tests run using the *anova* function in the order: day of inoculation: df = 3, $p < 10^{-15}$; base medium: df = 1, p = 0.0048; presence of dead cells: df = 1, p = 0.056).

4.3.3 Rescue in larger populations

One reason for a population to fail to undergo evolutionary rescue is that potentially adaptive mutations are rare and do not occur in that population. To help determine whether mutational opportunity was a limiting factor preventing evolutionary rescue in diploids, we conducted a mutant acquisition experiment with roughly 100-fold more initial cells (inoculation used 1 mL of overnight culture compared to 10 μL used in the main box acquisition experiments). Ten diploid populations, as well as one population of each haploid type, were exposed to 100 mL of YPDnystatin4 in 250 mL flasks. All populations grew within 10 days of inoculation, with the *MATa* and *MAT*α populations growing on Day 3, and the diploid populations growing on Days 8-10. Despite this growth, no diploid population grew when tested in YPDnystatin4 in the follow-up growth assay (Fig C.1). On the other hand, haploid populations generally showed reliable rescue as demonstrated by growth in follow-up assays, consistent with the smaller volume experiments. These results indicate that increasing initial population size roughly 100-fold was not sufficient to allow for beneficial mutations to occur in diploids during this short adaptation experiment.

4.4 Discussion

In this study, we have provided an example where ploidy alters the probability of evolutionary rescue. Diploids were generally not able to adapt to a high concentration of nystatin in the short time course given with none of 619 inoculated wells (excluding those found to have haploid contamination) showing evolutionary rescue, as compared with 216 out of 531 haploid wells. Although we expected rescue to be less

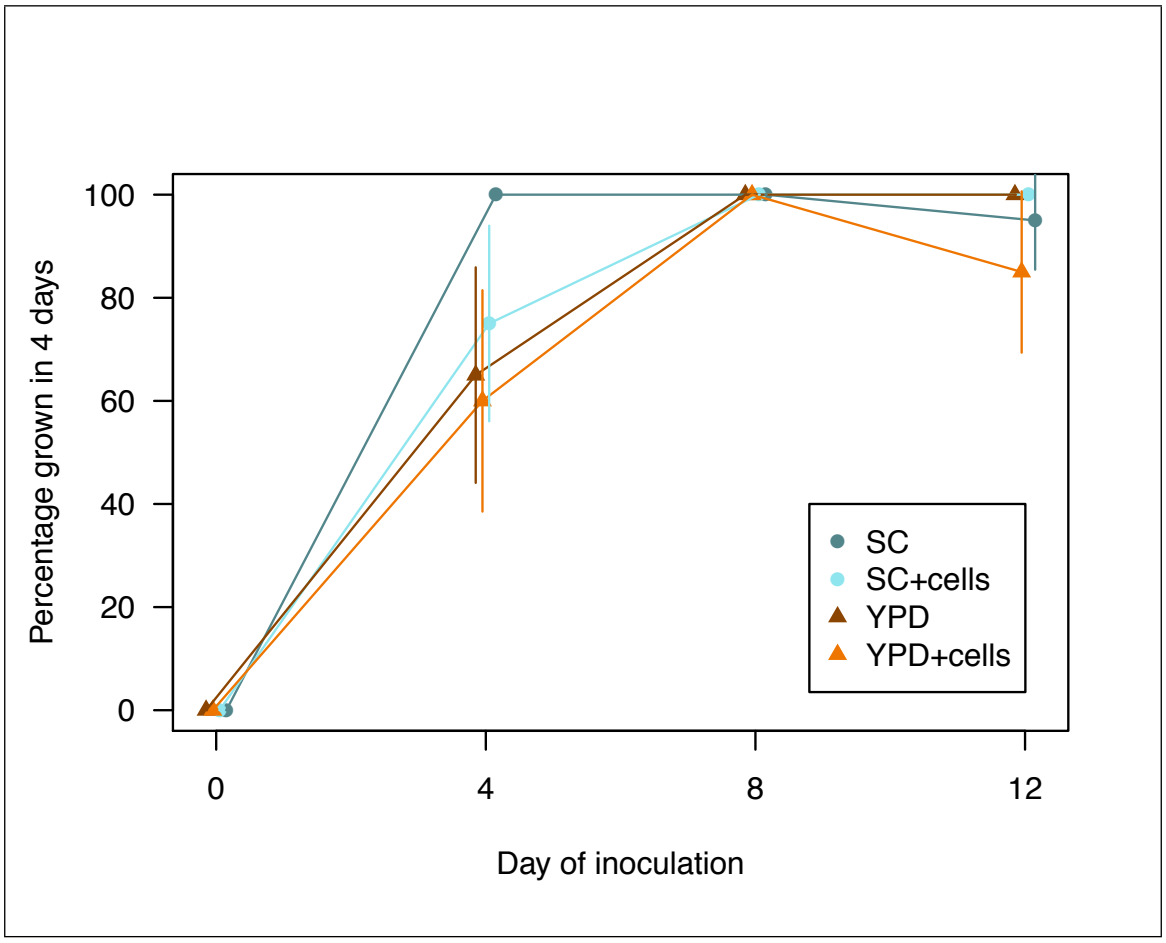

Figure 4.3: Plot of the percentage of wells inoculated on each day that showed growth within four days of inoculation in the nystatin efficacy experiment. Vertical bars represent 95% confidence intervals of the proportions. No wells grew within four days of the beginning of the experiment (when the nystatin was fresh). When four day-old nystatin was inoculated with yeast, many wells grew within four days, and all eight-day old nystatin wells grew within four days. The increased variability in growth observed among wells inoculated on Day 12 is likely due to the varying amounts of culture used to inoculate these wells (ranging from 2 μL to 10 μL). These results confirm that the nystatin may have been losing efficacy over the course of the initial acquisition experiments.

common in diploids, given previous work demonstrating that mutations accumulated in haploids were recessive (GERSTEIN *et al.* 2014), it was possible that diploids would explore beneficial mutations not observed in the haploids. We conclude that there is a genetic limit to adaptation in diploids with no simple one-step dominant rescue mutations available.

There are many differences between haploids and diploids that may cause the observed difference in their ability to undergo evolutionary rescue. Factors that may favour haploid rescue include larger population size, differing effect sizes of mutations depending on ploidy background (if mutations tend to be more beneficial in haploids), and low dominance of potentially beneficial mutations (OTTO and WHITTON 2000). These differences will be discussed in more detail below.

In yeast, haploids have smaller cells and therefore larger population sizes for the same volume of inoculum (MABLE 2001), with a less than two-fold difference. Larger population sizes in haploids should correspond to a larger number of mutations in these populations, but this is not the case because diploids have twice the genome (and therefore twice the number of mutational targets) when compared to haploids. To determine whether our starting population sizes were too small to allow diploid evolution, we exposed ten diploid populations to nystatin in flasks containing 100-fold more medium and initial inoculum. None of these populations underwent evolutionary rescue (Fig C.1). At this population size (roughly 7×10^7 cells), we expect 38% of sites to mutate within at least one of the flasks, with over 600 mutations per gene expected across the flasks, leading us to conclude that initial population size is not the problem (Section C.2).

The effect of potentially beneficial mutations may differ between haploids and diploids, even between haploids and homozygous diploids, which are often treated as equivalent. Previous work in this system has found that the effect sizes of beneficial mutations were not equal between haploids and homozygous diploids, for those mutations acquired in a haploid background, with haploids outperforming diploids (GERSTEIN 2013). In nystatin, these effect size differences may arise as a consequence of the geometrical differences between haploid and diploid cells. Nystatin acts by binding to ergosterol in the yeast membrane, making the membrane more permeable to ions, sugars and metabolites, resulting in cell death (CARRILLO-MUÑOZ *et al.* 2006). Because surface areas are higher for diploids than haploids (MABLE 2001), their sensitivity to nystatin may differ. In addition, the primary path to resistance to nystatin is through mutations in the ergosterol biosynthesis pathway, which affects the sterol composition of the cell membrane (GERSTEIN *et al.* 2012). Due to their larger surface area, diploids may suffer more of a fitness cost from these mutations due to decreased stability of the membrane, thus impacting the effective fitness benefit conferred. Nevertheless, ergosterol mutations still confer a large fitness benefit to diploids in nystatin when present in homozygous form (GERSTEIN 2013), suggesting that the difference in rescue probability between haploids and diploids is not due to differences in the effect of mutations when homozygous.

Another way in which the effect of potentially beneficial mutations can differ between ploidies is through their dominance. Many new mutations are recessive and therefore are only 'seen' by selection when rare in a haploid. Recessive beneficial alleles, especially those originating from new mutations, are unlikely to rescue a diploid population because they will not often spread to high enough frequency for homozygotes to be common (ORR and UNCKLESS 2008). Sex by random assortment is very unlikely to combine rare, potentially beneficial recessive alleles. Natural yeast perform a version of selfing wherein mating is most

common between gametes from a single diploid individual. This mechanism could produce an adapted, homozygous individual, but sex cannot be induced in our experimental setup, due to the short time frame and strong selection. Because reproduction is strictly asexual in our experiment, diploids must acquire a second mutation (either another new mutation or a loss-of-heterozygosity event) in order to gain any advantage from a recessive allele (MANDEGAR and OTTO 2007). Surprisingly, based on our results, we infer that there are no dominant or semi-dominant rescue mutations in this environment. This places a limit to evolutionary rescue on diploids at a lower concentration of the drug than for haploids. These results will not necessarily generalize to other environments because they depend on the nature of available adaptive mutations and the exact effects of the environment on the ancestral type.

ANDERSON *et al.* (2004) performed a similar experiment at high concentrations of another antifungal drug (64 and 128 μg/mL fluconazole), in which the initial lines were able to undergo seven to nine doublings but were unable to proliferate further. In their experiment, diploids evolved resistance more slowly than haploids but eventually all replicate diploid lines evolved heritable resistance (minimum inhibitory concentrations of 256 μg/mL fluconazole). Thus, diploids were not limited in their ability to undergo evolutionary rescue under their experimental conditions. Importantly, we have similar total numbers of cells in our inocula (accounting for the initial growth that was observed in fluconazole), so that the contrast between observing 100% rescue (ANDERSON *et al.* 2004) and our result of 0% rescue (for levels of drug used in the acquisition phase) must reflect a difference in limits to evolutionary rescue and not a difference in experimental power. While certain two-step mutations would rescue diploid populations in our experiment (such as mutations in both copies of an ergosterol pathway gene), they would require much larger populations than those used. We calculate a <0.0003 probability that a cell containing a secondary mutation in the same gene would be sampled across any of the deep wells or flasks in the experiment, assuming a relatively high rate of either secondary mutation or loss of heterozygosity (~10^{-4}) and assuming that any second mutation in the same ORF and/or mitotic recombination event would generate resistance, to be conservative (Section C.2). In contrast, ANDERSON *et al.* (2004) observed patterns consistent with such two-step mutations. The difference in observed two-step mutations likely reflects a difference in the selective environment. Because yeast undergo several generations in fluconazole before arresting growth, there is the opportunity for weakly resistant heterozygous mutations to increase in frequency in the populations, and therefore a higher probability of loss of heterozygosity for one of these mutations.

It may be the case that deterioration of the nystatin environment allows diploids to grow in our experiments sooner than they are able to adapt genetically. Phenotypic heterogeneity (without an underlying genetic basis) in the ability to persist in the presence of antibiotics has been observed in bacterial populations (e.g., BALABAN *et al.* 2004). These "persister" types remain sensitive to the antibiotic, however, when retested. Such persistence could explain the presence of nystatin-sensitive populations among those that grew in the initial acquisition experiments. The populations may persist at low numbers while the concentration of nystatin is high enough to be inhibitory and then show growth once the efficacy of nystatin has dropped below some threshold. This is consistent with the observation that diploids tend to grow on the later days of the acquisition experiments, which is when we also observe growth of non-resistant haploid populations (Fig. C.2). Follow up experiments evaluating the efficacy of nystatin over time indicate that nystatin likely

lost efficacy by this time (Fig. 4.3), allowing the growth of lower tolerance strains. While we conclude that full resistance to 4 μM nystatin was not exhibited by any of the diploid lines assayed, it remains possible that the diploids did evolve low levels of resistance that improved their ability to persist or to grow once nystatin became less effective.

In our study, we appear to have exceeded the limit of genetic adaptation possible in diploids, but not in haploids, by using a high concentration of the fungicide nystatin and a short time frame. A previous study found that diploids were able to adapt at the same rate as haploids to a lower concentration of nystatin (0.6 μM) over a longer period of time (140 generations) (GERSTEIN *et al.* 2011) under conditions that allowed growth of the initial strains (i.e., not an evolutionary rescue experiment). It is possible that the larger number of generations in that experiment provided the opportunity for strains to get the kinds of two-step mutations that seem to be necessary for resistance to high concentrations of the drug. However, initial whole-genome sequence data from these strains found no mutations in either *ERG3* or *ERG6* (data not shown), the most commonly used genes in haploids at high concentrations of nystatin (GERSTEIN *et al.* 2012). Instead, the ability of the diploids to evolve in GERSTEIN *et al.* (2011) suggests that different, and potentially more dominant, mutations may be available at lower concentrations of nystatin that are not sufficient to provide resistance to higher concentrations.

We find that evolutionary rescue is not always possible and that the limits can depend strongly on the ploidy of the organism in question. These results have implications for conservation. For example, among algae, we might expect evolutionary rescue in the face of climate change to change the relative proportions of species with a haploid phase (haplonts or haploid-diploid species) relative to those with only a diploid phase (diplonts). As another example, higher standards may be needed for pollutants/toxins that require exposed organisms to adapt using recessive mutations, because of the risk that evolutionary rescue will fail. There are also implications for disease management. By investigating the genetic basis of potential resistance to our treatments of choice (antibiotics, pesticides), we can make informed decisions about timing and dosage. In this way, we can endeavour to minimize potentially harmful evolutionary rescue in the organisms that we are attempting to control.

Chapter 5

Discussion

In this thesis, I have broadly addressed questions about the genetics of adaptation and speciation using the yeast *Saccharomyces cerevisiae* as a model system. I have used the tools of experimental design and statistics along with laboratory techniques for yeast manipulation and assessment to investigate the genetic basis of evolution, taking advantage of the fast generation time of yeast and its genetic tractability. By allowing the organism to explore possible mutations naturally and then sequencing to discover the utilized mutations, we can begin to understand why evolution proceeded the way that it did. The use of experimental evolution greatly facilitated this process, as studies of natural systems are often limited by the precision with which they can map causative mutations, usually only able to narrow down the genome to large blocks of interest. For many species, these studies can often be restricted by either having relatively little knowledge available about the underlying genes or only having the ability to investigate candidate genes. I found that the genetic details of evolution are sometimes surprising, and they force us to expand our thinking about how evolution 'typically' proceeds. Theoretical models of evolution must, by their nature, make certain assumptions about the underlying genetic system, and I hope that results from these experiments will inform theoreticians about interesting aspects of genetics that warrant further investigation. The chapters of my thesis have investigated the repeatability of adaptation in experimental strains of yeast adapted to high concentrations of copper (Chapter 2), tested genetic interactions between first-step adaptive mutations (Chapter 3), and explored the limits of genetic adaptation of haploid and diploid yeast in a fungicide (Chapter 4). I will discuss the main conclusions of these chapters in turn and then use Fisher's geometric model as a theoretical framework to revisit some of the data from two of these chapters.

5.1 Thesis summary

5.1.1 Chapter 2: Repeatability of adaptation

In Chapter 2, I explore the repeatability of adaptation to high concentrations of copper. I find that the level of repeatability depends on what one defines as repeatable. The same gene (*CUP1*) was involved in increases in copy number in almost all replicates (27/34), as expected from previous work on copper resistance in yeast (ADAMO *et al.* 2012; FOGEL and WELCH 1982; FOGEL *et al.* 1983), so we might conclude that evolution was highly repeatable. The mechanism of copy number increase varied between strains, however, with some strains utilizing whole-chromosome aneuploidy (an additional copy of the relevant chromosome) and others increasing copy number of the already tandemly repeated region. Thus, at the level of mutation type, adaptation was slightly less repeatable. Finally, many strains also carried secondary genic mutations. Of the affected genes, some were mutated in a few different strains (highly unlikely by chance alone), while some

genes were mutant in only one. These genes had a variety of functions and cellular localizations. When also considering these genic mutations, the level of repeatability was quite low. These results highlight an important contribution of experimental evolution to evolutionary theory. Observations like this force us to consider what is meant by the term 'repeated', both in terms of the gene involved and the mutation type, highlighting the importance of the varying genetic mechanisms available to evolution.

By comparing the results of Chapter 2 with those of GERSTEIN *et al.* (2012), which uses the same experimental setup and same starting yeast strain but a different selective environment (a high concentration of the fungicide nystatin), we can observe how genomic breadth, and therefore repeatability, changes with the agent of selection. Few studies have directly compared the repeatability of adaptation between environments in this way (but see GRESHAM *et al.* 2008), but it is difficult to draw conclusions about the effects of a single factor (environment) when comparing studies that have been performed with many other varying factors (e.g., population size, population dynamics, organism of study). In our comparison, I find a variety of mutation types in copper (copy number variation, aneuploidy, genic mutations) while GERSTEIN *et al.* (2012) only find genic mutations in nystatin. Considering the biological pathways targetted, evolution in copper involved many different pathways and biological functions, while adaptation to nystatin was acquired through mutations in genes of a single biosynthetic pathway. Finally, at the level of individual genes implicated in adaptation, increased copy number of a single gene, *CUP1*, was almost uniformly observed in evolution in copper. It is likely that amplification of this locus had a high mutation rate because it is already present in a tandemly repeated region, which is prone to unequal crossover, gene conversion or single-strand annealing (ZHANG *et al.* 2013). Other genes were repeatedly involved in copper adaptation, but much less frequently. In contrast, adaptation to nystatin involved one of four genes, and most adapted strains carried mutations in one of the two repeatedly used genes (*ERG3* or *ERG6*). Overall, adaptation to nystatin was more repeatable than adaptation to copper at both the level of mutation type and biological pathway. At the level of individual gene, adaptation to copper was simultaneously more repeatable, due to the almost uniform observation of *CUP1* amplification, and less repeatable, due to the observation of a variety of other genic mutations involved in multiple cellular processes.

Unfortunately, due to the high mutability of the tandemly repeated *CUP1* region, it was difficult to determine the effects of genic mutations that co-occurred with high *CUP1* copy numbers since lines with high *CUP1* copy number were unstable (did not segregate 2:2 in cross progeny). For repeatedly hit genes, the benefit of these mutations was supported by many lines of evidence, including their presence in multiple independently adapted strains, but it would have been much more difficult to characterize all of the other genic mutations observed, and this was not done. Most studies that focus on genetic repeatability mainly refer to repeatably used genes. Going forward, it would be insightful to characterize the rarely-used genes in evolution to determine whether they are rare because the mutations have small fitness benefits or because the mutations themselves are rare, or whether it is often a combination of the two.

5.1.2 Chapter 3: Evolution of BDM incompatibilities between first-step adaptive mutations

In Chapter 3, I investigate epistasis between first-step adaptive mutations in the fungicide nystatin. Epistasis between beneficial mutations from the same environment can have implications for evolutionary trajectories,

affecting both the speed and direction of evolutionary change, and is relevant for speciation. Independently evolving populations may adapt to similar selection pressures via different genetic changes, even if they are exposed to an identical environmental challenge. How these genetic changes interact in hybrids between the populations may determine whether or not the populations evolve to become new species. I used mutations from an initial study of haploid adaptation to nystatin (GERSTEIN *et al.* 2012), focusing on one mutation in each of four genes in the ergosterol biosynthesis pathway. I found that genetic interactions were prevalent and predominantly negative, with the majority of mutations causing lower growth when combined in a double mutant than when alone in a single mutant and, in one third of cases, the growth of the double mutant was lower than either single mutant. Thus, BDM incompatibilities evolved readily, even among populations adapting to identical conditions. The prevalence of these kinds of interactions is surprising given the small number of mutations tested and demonstrated that postzygotic reproductive isolation could evolve between populations differing by only a single genetic change each. These results lend support to the mutation-order model of speciation where populations accumulate reproductive isolation due to the chance order in which mutations arise in each. This model is difficult to prove in nature due to the requirement of known parallel selection histories and has not been thoroughly investigated in the lab. Further, the observation of isolating epistasis between first-step mutations may drive the evolution of further divergence if one considers that these early mutations will potentially constrain future evolutionary paths to become increasingly incompatible.

The observation of sign epistasis runs counter to expectations for mutations arising in a single biosynthetic pathway in the face of a simple selective pressure. We would expect these mutations to mask each other (the upstream mutation masking the effects of the downstream one) (AVERY and WASSERMAN 1992). Similarly, if we use metabolic network theory to predict the epistasis between different genes in a linear pathway (SZATHMÁRY 1993), we would expect that if two genes independently reduce flux through a pathway then the double mutant should reduce flux more than either one alone but less than the combined effects (diminishing returns epistasis, not sign epistasis). The prevalence of BDM incompatibility-type interactions indicates that the situation is not simply described by the linear biosynthetic pathway. Along these lines, we also found that the phenotype of the double mutants was not reflected by the pathway position of the single mutations, with sterol profiles of double mutants often matching one of the two single mutants but not always the upstream one. The sterol profile relationships did not match the fitness relationships either, implying that genes are deviating from the described pathway and that the mutations are changing something about the cell other than just the sterol profile in order to affect fitness.

The nature of the genetic interactions depended not only on the mutations involved but also the concentration of drug in the assay conditions. When two strongly beneficial mutations were combined, I found that all double mutants had equivalent or lower fitness than the two parent single mutants in a non-stressful or mildly-stressful environment, resulting in negative genetic interactions. When the stress (concentration of nystatin) was increased, performance of the double mutants reversed; they were often the most fit strain and, at very high drug concentrations, were the only ones able to survive and grow, resulting in very positive genetic interactions. This result is not initially intuitive but has theoretical grounding in Fisher's geometric model of adaptation, as will be explored in Section 5.2.1, below. This was an especially interesting finding to make as I am not aware of many other similar examples, especially among BDM incompatibility-type

interactors, and demonstrates the sensitivity of reproductive isolation to the environment in which hybrids are formed.

The scope of this study was limited, however, and investigated only six pairwise combinations of alleles. In addition, these results may be specific to the nature of the mutations involved. The prevalence of negative epistasis might come from the fact that all mutations investigated were in the ergosterol biosynthesis pathway. Negative epistasis can be observed as a result of two partial loss of function mutations in an essential pathway if each decreases flux through the pathway (BOONE *et al.* 2007), which could explain our observations because ergosterol is the main sterol in the yeast membrane and is therefore ostensibly essential. In order to determine the generality of these conclusions, similar studies should be performed for a larger set of beneficial mutations across a variety of environmental conditions.

5.1.3 Chapter 4: Limits to adaptation

In Chapter 4, I compare adaptation of haploids and diploids in the fungicide nystatin. I set out to investigate the different genetic paths taken by haploid and diploid yeast evolving in a concentration of nystatin that inhibits growth. As described briefly above, previous work in the lab (GERSTEIN *et al.* 2012) found that haploid strains adapt to this environment by acquiring mutations in one of four genes of the ergosterol biosynthesis pathway. To determine how diploid evolution might differ, we repeated the experiment using haploids of both mating types and diploids. There are many reasons why adaptation might proceed differently in haploid and diploid yeast including differences in number, distribution and types of mutations (due to genome size), dominance of adaptive mutations, fewer diploid cells in the same volume of culture due to cell size (MABLE 2001), and the same mutations having different effects or effect sizes in haploids vs. homozygous diploids (as found for some nystatin-resistance mutations by GERSTEIN 2013).

Instead, I appear to have found the limit for diploid genetic adaptation through one-step mutations. We found no cases in which a diploid population evolved full resistance to the original evolutionary conditions. I do not believe that sampling effort was limiting in this case. We found no diploid mutants after testing over 600 wells, compared with finding 216 haploid mutants out of 531 wells tested. In addition, I did not find a diploid mutant in any of the ten flasks tested, compared with both haploid flasks exhibiting genetic rescue. It seems that diploid yeast are not able to obtain single mutations of large enough effect to allow growth. There is previous evidence that diploids are able to adapt at the same rate as haploids to a lower concentration of nystatin (0.6 μM, which does not fully inhibit growth), over a longer time course (GERSTEIN *et al.* 2011), but the same is not true at this higher concentration (4 μM). All adaptive mutations previously found in this inhibitory concentration of nystatin in haploids were recessive (GERSTEIN *et al.* 2014). Likely, the only single-step mutations that have a large enough effect to genetically rescue a population in this environment are recessive, and two mutational events at a single resistance-conferring gene are too rare to rescue the tested populations. This chapter provides an example of how the adaptive limit of an organism can depend on both the exact environment in question (concentration of nystatin) and the biology of the organism (ploidy), having implications for the role of ploidy in evolutionary rescue in the face of a strong selective pressure such as an antibiotic.

Interpretation of the results of this experiment was difficult because, in many instances, growth was

observed without the evolution of heritable resistance to the original environmental conditions. In order to isolate first-step, large-effect mutations, we used a concentration of nystatin that inhibited the growth of wild-type yeast. In this environment, if yeast were to be serially transferred into fresh medium, they may be driven to extinction by dilution before being given the opportunity to adapt. That is why we did not transfer yeast over the course of the experiment but instead allowed them to remain in the original medium until growth was observed. Unfortunately, because nystatin is not stable over long periods of time in 30°C, the efficacy of the drug decreased over the course of the experiment, leading to the growth of many non-resistant or partially-resistant yeast populations. Deciding on a cutoff for evolutionary rescue was therefore difficult, because some populations appear to be at their borderline for growth under the test conditions. In addition, we could not allow diploids more time to adapt because the environment was not stable for longer periods of time. I would suggest that this experiment be repeated in other environments that are more likely to be stable for extended periods of time, or using careful addition of drugs over time, to test the generality of these conclusions.

5.2 Putting adaptive mutations in the context of Fisher's geometric model over environmental gradients

In this section, I would like to revisit some of the results from Chapter 2 and Chapter 3 in the context of Fisher's geometric model (FISHER 1930). Fisher's geometric model has been used to describe theoretical reasons for epistasis between adaptive mutations (BLANQUART *et al.* 2014) and for patterns observed in speciation (FRAÏSSE *et al.* 2016). It has also been combined with experimental work to describe patterns in the mutations that arise in different concentrations of an antibiotic and how the optimum shifts between those concentrations in *Escherichia coli* (HARMAND *et al.* 2017), as well as to test diminishing-returns epistasis in a multicellular fungus, *Aspergillus nidulans* (SCHOUSTRA *et al.* 2016). I will take a less formal approach than the above-mentioned papers and use Fisher's geometric model as a framework to think about patterns of epistasis and how the fitness of mutants might change in different levels of the same stressor.

In Fisher's geometric model, the fitness of a genotype is determined by its distance in phenotype space from the optimal phenotype for that environment as well as a function that describes how fitness drops off with increasing distance from the optimum. Let's imagine a population starting some distance from the fitness optimum in an environment. If it gains a beneficial mutation, we can imagine the mutation as a vector in n-dimensional phenotype space that brings the population closer to the fitness optimum. This mutation fixes and now, when a second mutation arises, it is determined to be beneficial if it again brings the population closer to the fitness optimum. An important point to consider, however, is that this mutation would not necessarily have been beneficial in the original genetic background, depending on the direction and size of the first mutation vector. For the same reason, two independently adaptive mutations (both beneficial in the original genetic background) may not be compatible with each other if, when combined, they cause the organism to overshoot the optimum (adding both vectors together results in a phenotype that is further from the fitness optimum than either vector alone, Fig. 5.1). This is one example of how epistasis for fitness can arise in this model despite additivity of the underlying mutations and phenotypes,

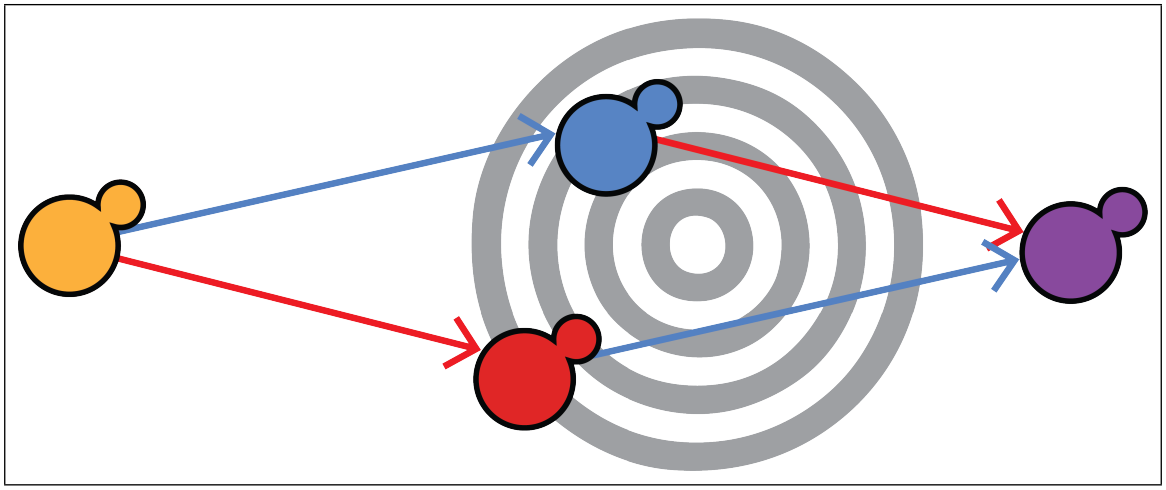

Figure 5.1: Optimum overshooting (sign epistasis) in Fisher's geometric model. Using Fisher's geometric model of adaptation (FISHER 1930), we can imagine two genotypes (red and blue yeast) that are adapting from a common ancestor (yellow yeast) to one optimum (the centre of the grey target). The initial beneficial mutations (red and blue arrows), represented as vectors in 2-dimensional space, put the genotypes in different locations in phenotype space. Later mutations will be beneficial or deleterious depending on the genetic background in which they arise and whether they bring the genotype closer to the fitness optimum. The two initial beneficial mutations are mutually exclusive, not conferring a fitness benefit in each other's background, resulting in a less fit hybrid combination (purple yeast) that overshoots the phenotypic optimum.

and we observe the outcome as sign epistasis in fitness between independently adaptive mutations. In this way, evolutionary trajectories can be constrained by the size and direction of mutation vectors in phenotype space, and reproductive isolation due to genetic incompatibility can arise between independently evolving populations.

I will attempt to describe two datasets using a simple form of Fisher's geometric model. First, I will investigate the single and double mutant fitness data in different concentrations of nystatin from Chapter 3, to see whether the patterns of epistasis for fitness are consistent with underlying phenotypic additivity. Second, I will use the data measuring tolerance to copper from Chapter 2 to see whether fitness in a range of concentrations of copper can be predicted from a strain's fitness in a single concentration of copper given information about how other strains react across the range. Both of these analyses will utilize data from a set of genotypes measured in a series of concentrations of the stressor to which they have adapted. In Fisher's geometric model, changing environments are described by changing optima, and there are two main ways in which I will discuss an optimum changing (Fig. 5.2). The first is a shifting of the optimum as environments change. In this case, the optimum shifts in multidimensional phenotypic space, making different phenotypes either beneficial or deleterious depending on their location in that space. The second is a narrowing (or widening) of the optimum. In this case, the most fit phenotypes will remain the most fit, but relative and absolute fitness values change.

In order to see how well Fisher's geometric model describes the observed data, we will need to place our mutants in phenotype space. We will use fitness data for each strain, preferably in their adaptive environment,

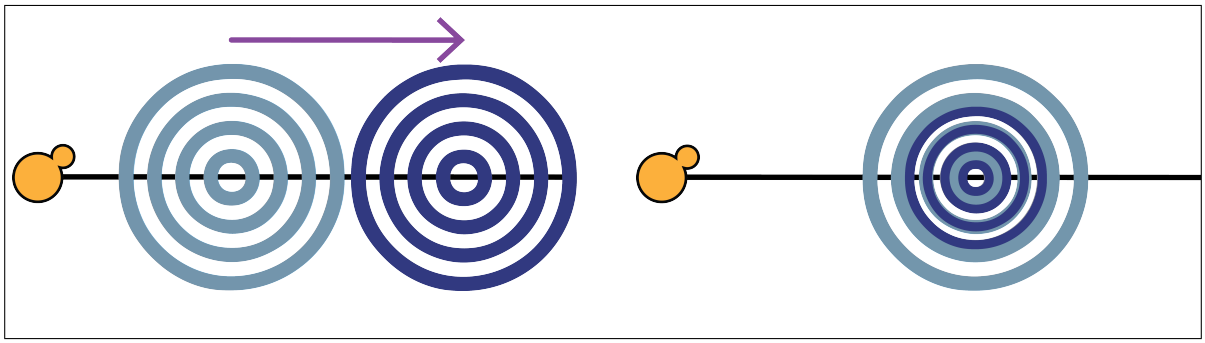

Figure 5.2: Optima change with changing environments. In Fisher's geometric model, different environments are described by different positions and distributions of the phenotypic optimum. There are two main ways in which I might expect the optimum to change. First, as illustrated on the left, the optimum might shift in location in different environmental conditions. Second, as illustrated on the right, the optimum might narrow (or widen) with changing conditions, making certain genotypes more or less fit without changing their distance to the optimum. In both cases, the initial fitness surface is represented by a light blue target and the dark blue target is the fitness surface after an environmental change.

to determine the relative positions of the strains. Imagine a one-dimensional line drawn between the ancestor and the fitness optimum. All first-step adaptive mutations will be vectors that originate from the ancestor, but they may point in any number of directions and cover any amount of distance, as long as they end closer to the optimum. However, the projection of those (potentially multi-dimensional) vectors onto the one-dimensional line drawn between the ancestor and the optimum will determine the fitness benefit conferred by each mutation (Fig. 5.3). I will therefore use differences in fitness measurements as proxies for distance from the ancestor towards the optimum in one-dimensional space. While it is possible to overshoot the optimum (while still ending closer to the optimum than where the ancestor started), I will assume that this has not occurred in the adaptive environment. This assumption is supported by the fact that the ancestor cannot grow in the selective environment and is therefore likely to be initially far from the optimum. I will also assume linear mapping of fitness differences onto distance travelled in this one dimension, which does not have to be the case. I will use this technique to map mutants onto phenotype space. I will then keep phenotypic distances between strains constant as I investigate what occurs when the concentration of the stressor is changed.

5.2.1 Can changing epistasis be explained by a shifting optimum? Revisiting Chapter 3

In Chapter 3, we observed changing epistasis for optical density (OD) after 24 hours of growth with changing concentrations of nystatin. When no nystatin was present in the medium, the ancestor was most fit but all strains grew well. When low to moderate concentrations of nystatin were present, the single and double mutants all had relatively high growth, indicating negative epistasis because double mutants were less fit than expected. Finally, as the concentration of nystatin was increased, double mutants were often the only strains capable of growth. These observations led us to conclude that incompatibilities (as determined by sign epistasis) for growth rate that were present at moderate concentrations of nystatin (2 μM) were not stable

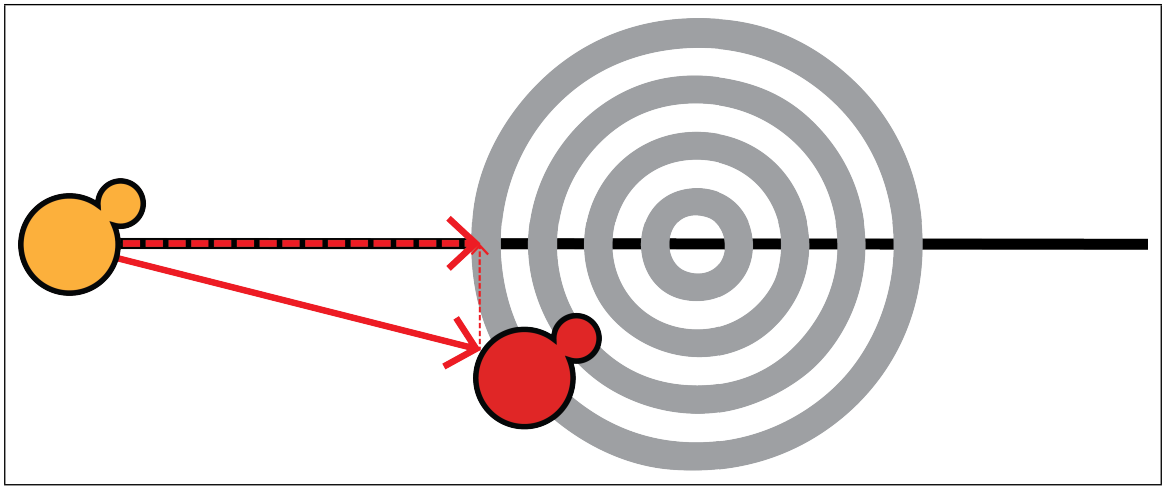

Figure 5.3: Projecting mutational effects into one dimension. If we were to draw a straight line connecting the ancestral genotype (yellow yeast) and the fitness optimum (centre of the grey target), we could then project all of the mutational effects onto this dimension. This is shown for the red mutation (solid line). By projecting the red mutation onto the black line by way of the small dashed arrow, we can obtain the distance travelled by that mutation in only the dimension of interest (red dashed arrow). The distance travelled along this one dimension is the relevant distance for fitness benefits conferred by the mutation.

across different concentrations of the drug.

I have replotted the OD data using the method described in the previous section in order to determine whether these fitness measurements are consistent with underlying additivity at the phenotypic level and a shifting optimum with changing concentrations of the drug. First, to place the single mutants along the x-axis dimension, the difference in OD after 24 hours of growth between the single mutants and the ancestor was found for 4 μM nystatin, the adaptive environment. Then, I placed the double mutant strains a distance from the ancestor that was the sum of the distances from the ancestor of the two single mutants of which the double mutant is composed. For example, the *geneX geneY* double mutant would be found at a distance of 10 'units' from the ancestor if *geneX* were 2 'units' from the ancestor and *geneY* were 8 'units' from the ancestor. Once strains were positioned along the x-axis, I plotted the OD after 24 hours of growth for all concentrations of nystatin (Fig. 5.4).

When we look at the data, we notice that the method of placing strains along a single dimension works relatively well to produce interpretable results under Fisher's geometric model. For each concentration of nystatin, we can imagine a single fitness peak, and the fitness of the strains drops off roughly consistently with their distance from this peak along the x-axis. Because this remains true for each concentration of nystatin without changing the positions of the strains along the x-axis, the data is broadly consistent with a peak changing in this one dimensional space. In contrast, we might have found that the strains needed to change relative positions along the x-axis in order to find a single fitness peak, which would be interpreted as the relevant one-dimensional axis itself shifting in phenotype space (if the target in Fig. 5.3 were to move off of the black line, the projection of the red mutation onto the black line would no longer be useful for ordering mutations relative to the optimum). There are two main implications of the consistency of the data

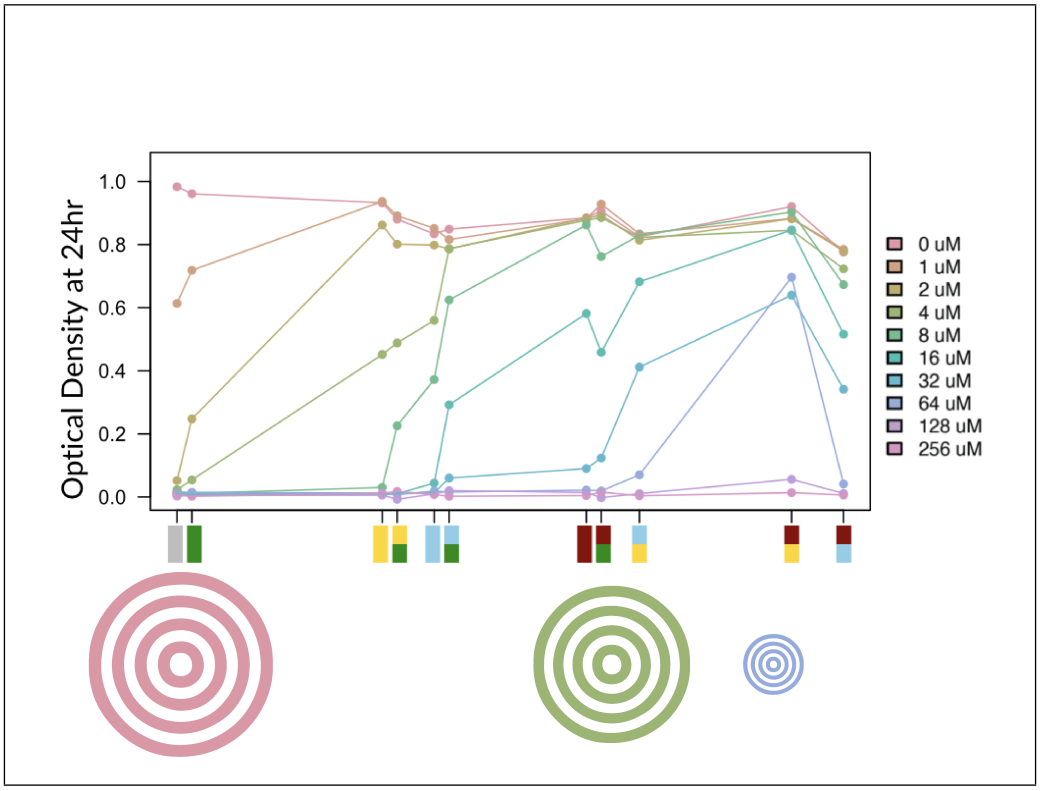

Figure 5.4: Changing epistasis in OD after 24 hours of growth in an environmental gradient. Mean OD after 24 hours of growth in different concentrations of nystatin is plotted by strain. First-step adaptive mutations in nystatin are plotted along the x-axis according to their proportional fitness benefit in 4 μM nystatin (the original evolutionary environment, olive green) when compared with the ancestor (grey box). Double mutant strains are plotted a distance from the ancestor equal to the sum of the distance travelled by each of their composite single mutations along the x-axis. Strains are represented by coloured boxes indicating their mutations, with single mutants represented by a single coloured box (green: *erg5*, yellow: *erg3*, blue: *erg7* and red: *erg6*) and double mutants represented by two coloured boxes stacked together. Each concentration of nystatin is plotted with a unique colour (see key on the right). On the bottom is a visual representation of how the optimum may be changing in different concentrations of nystatin. Only three optima are shown, for simplicity, and each is coloured according to the concentration that they represent. Both shifting and narrowing are represented as the concentration of nystatin increases.

with Fisher's geometric model. First, our results are broadly consistent with an underlying additive basis for double mutant phenotypes. The phenotypic trait could arise from both single mutations decreasing flux through the ergosterol biosynthesis pathway. We would expect two flux-reducing mutations to further reduce flux when combined as a double mutant, and this relationship could be additive on the one-dimensional phenotype axis. Second, the changes in fitness of the different strains as the concentration of nystatin increases seem to result from a combination of both shifting and narrowing of the fitness optimum. At very high concentrations of nystatin (e.g., 64 μM), narrowing of the optimum is most obviously observed centred near the *erg3 erg6* double mutant strain (yellow and red box in Fig. 5.4), which retains its ability to grow despite all other strains failing to do so. Shifting of the optimum is observed as only strains on the right-hand side of the plot maintaining their ability to grow in moderate to high concentrations of nystatin. If we again focus on the *erg3 erg6* double mutant, despite its high fitness at high concentrations of nystatin, there is no evidence of this same strain being the most fit at low and moderate concentrations of nystatin, which is consistent with an optimum that is shifting. The epistatic relationships between mutations arise as a consequence of the combination of additivity on the phenotype axis (x-axis) and the position of the fitness optimum.

From the plotted data, it seems likely that the sign epistasis observed for growth rate in 2 μM nystatin was due to overshooting of the optimum in double mutant strains. This overshooting is difficult to observe from the OD data in 2 μM nystatin, but growth rate is probably a more sensitive measure of fitness, which allows better discrimination of fitness differences. Especially at lower concentrations of nystatin, where most strains are still capable of some growth, by waiting 24 hours until taking the OD measurement, we are leaving time for slower strains to 'catch up' in growth. Unfortunately, due to space constraints, we were not able to perform growth rate analyses in a large number of concentrations of nystatin (the plotted data is underlain by 3,840 measured wells, while only 400 wells can be simultaneously measured for growth rate). Similarly, the growth rate data presented in Chapter 3 includes more replicates per strain than this OD data, so we have greater confidence in each data point.

5.2.2 Can the fitness of a strain in an environmental gradient be predicted from its fitness in a single environment? Revisiting Chapter 2

Inspired by the results of Section 5.2.1, I wanted to find out whether we could generally order adapted strains along a one-dimensional phenotypic axis by knowing their evolutionary environment and determine how the optimum phenotype changes over an environmental gradient. If this were possible, then we could theoretically place newly discovered strains along that axis based on their fitness in one environment, and potentially infer their relative fitness in the rest of the gradient. In order to determine whether this would be plausible, I revisited the dataset from Chapter 2 where copper tolerance was measured for all adapted 'Copper Beneficial Mutation' (CBM) lines. In this chapter, tolerance was calculated as the inhibitory concentration 50 (IC50), which was determined by the OD of all strains after 72 hours of growth over a range of concentrations of copper.

I plotted the underlying OD data as described above with one important exception. I could not use the original evolutionary environment (12.5 mM copper) to order the strains because many strains did not have higher fitness than the ancestor in this environment. All strains have increased copper tolerance relative to

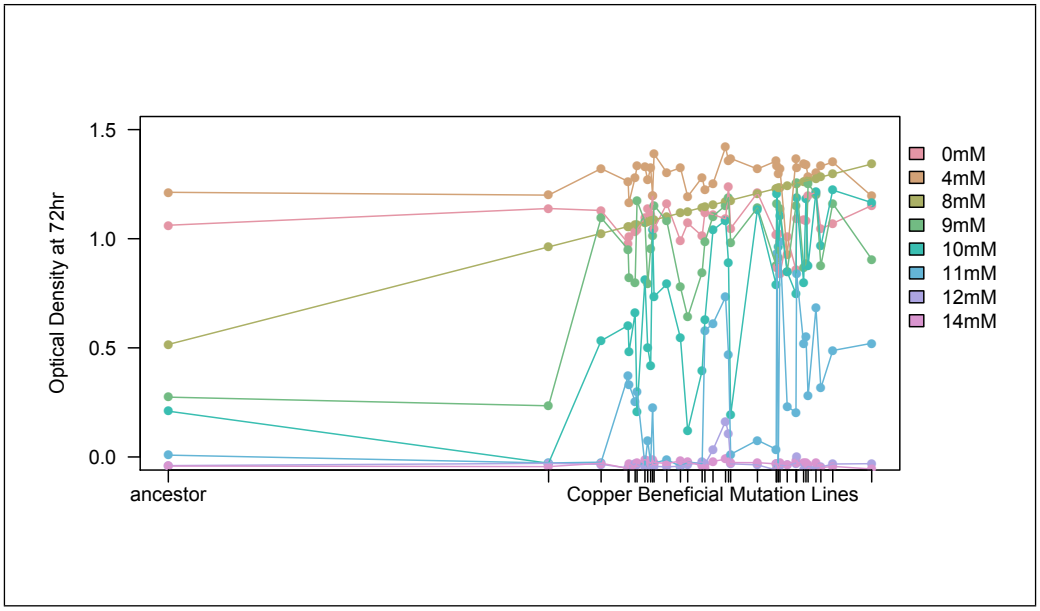

Figure 5.5: OD of copper beneficial mutation lines after 72 hours of growth in a gradient of copper concentrations. Mean OD measured after 72 hours of growth in different concentrations of copper is plotted by strain. CBM lines are plotted along the x-axis according to their proportional fitness benefit in 8 mM copper (olive green) when compared with the ancestor (left hand side). Note that the original evolutionary environment (12.5 mM copper) could not be used in this case because many strains did not have a higher OD than the ancestor in this environment. 8 mM was chosen as the environment closest to the ancestor's IC50. Each concentration of copper is represented by a unique colour (see key on the right).

the ancestor (Fig. 2.2), but none have tolerance to 12.5 mM when measured after 72 hours. Their ability to grow in the initial acquisition experiment under these conditions was likely due to a prolonged period of sequestration of copper by tolerant cells, which might have decreased the effective concentration of copper in the medium and allowed for rapid proliferation. Indeed, the first copper-adapted lines were only isolated seven days after inoculation, which is much later than the similarly acquired nystatin-adapted lines from GERSTEIN *et al.* (2012) that were all isolated within seven days of incubation. Because I could not use the data from 12.5 mM copper to order the strains, I instead used the data from 8 mM copper, as this concentration was the closest to the ancestor's IC50 (Fig. 2.2). The results are plotted in Fig. 5.5.

Based on Fig. 5.5, I would expect the predictive power of this dataset to be quite low. There is no clear single-peaked optimum at most concentrations of copper, and it does not shift or narrow in any broadly observable way. There are many reasons why this might be the case. First, as mentioned above, I could not use the adaptive environment to order the mutants. By using a lower concentration of copper, I might have chosen an environment where the actual optimum is somewhere intermediate compared to the true distribution of the strains, thus breaking the assumption that we can determine the length of each mutation along the single dimension by the difference in fitness between that strain and the ancestor. In fact, even if we did know the 'true' underlying fitness optimum in the original adaptive environment, we may have strains that have overshot the optimum (while still remaining fitter than the ancestor) in the original set. Many more

mutations were observed per strain in copper than in nystatin, and the nature of the *CUP1* mutation is such that it may actually change over time in the strain while we are measuring it (placing progeny of a single test at different positions along the axis). Additionally, our assumption that one phenotypic dimension should be able to order mutants in relation to the optimum may not hold for this set of lines as there are many more types of mutations (copy number variation in *CUP1*, chromosomal aneuploidy) in addition to genic mutations among these strains, and many more biological processes involved. The complexity observed may necessitate the use of multiple dimensions to properly map the mutations in phenotype space.

The nature of the selective environment may also be an important difference between the two datasets explored. While nystatin only has negative effects on the cells, copper does not. Cells require some amount of copper to survive. In fact, strains perform better in 4 mM copper than 0 mM (including the ancestor; note that this is copper addition on top of what is normally present in rich YPD medium) and many strains are even more fit in higher concentrations (compare Fig. 5.5 with Fig. 5.4 where growth is generally best in 0 μM nystatin, and does not show large improvements in higher concentrations). The main mutation conferring copper tolerance is amplification of *CUP1*, which is a metallothionein protein that binds copper. It is possible that having many copies of *CUP1* in the cell is resulting in too many copper ions being sequestered away in lower concentrations of copper, but allowing for better utilization of copper in high concentrations (even better than in 'non-stressful' conditions). The complicated relationship between environmental concentration and fitness is making fitness unpredictable in the one-dimensional case. We might have success by partitioning mutations along two axes, depending on whether or not they are in the *CUP1* locus (similar to what was done in HARMAND *et al.* 2017 for *gyrA* and *non-gyrA* mutants). Finally, because OD was measured after 72 hours of growth in this assay, it is possible that early fitness benefits/deficits in these strains that would be better described by Fisher's geometric model are being masked by later growth. The lack of generality of this method of placing mutant strains in phenotype space that can be translated into fitness by a singular phenotypic optimum indicates that there is still much to learn about the relationships between genotype, phenotype and fitness and how they are affected by evolution.

5.3 Conclusions

My work has implications for understanding of the genetic basis of adaptation in different types of environments, levels of the same environment, and genetic backgrounds or ploidies. I find that the genetic repeatability of adaptation depends on the genomic target size in the adaptive environment and that this target size can also be influenced by the ploidy of the organism. I also investigate genetic interactions between adaptive mutations and find that different mutations adapting populations to the same environment might often lead to BDM-type incompatibilities, with consequences for the likelihood of mutation-order speciation. Finally, I find that the nature of those genetic interactions depends on the environment in which they are measured, with negative interactions in the fungicide nystatin becoming positive in higher concentrations of the drug. The observed changes in interactions are consistent with a shifting and narrowing fitness optimum under Fisher's geometric model in a one-dimensional phenotype space. When attempting to similarly place copper-adapted strains in a one-dimensional phenotype space, I found that simple fitness optima

were not readily apparent. The usage of Fisher's geometric model for predicting the changes in fitness of adapted strains over environmental gradients may be limited by the biological nature of the interaction of the organism with the environment in question, as mediated by the adaptive mutations.

I believe that future investigations into the links between genotype, phenotype and fitness are important for progressing our models of evolution. Determining how mutations can lead to specific changes to phenotype and fitness may depend on understanding these links at the molecular or biochemical level. We have made great strides in our ability to broadly predict genes involved in adaptation (e.g., ergosterol pathway genes in nystatin resistance and *CUP1* in copper resistance), but our ability to predict specific aspects of adaptation (e.g., precise fitness effects of mutations, the type of epistasis between them, and their sensitivity to concentrations of a stressor) remains poor. By integrating molecular and biochemical information about adaptive mutations, including whether they act in biochemical pathways or protein complexes and their regulatory relationships, we may improve our ability to map genotype onto phenotype. Generalities in these principles would inform theoretical models of evolution, such as Fisher's geometric model. As our ability to determine the genetic basis of evolution grows with the increasing ease of genomic studies, the next step is to improve our mechanistic understanding of how those genetic changes lead to changes in fitness.

Bibliography

ABRAMOFF, M., P. MAGALHAES, and S. RAM, 2004 Image processing with ImageJ. Biophotonics International **11**: 36–42.

ADAMO, G. M., S. BROCCA, S. PASSOLUNGHI, B. SALVATO, and M. LOTTI, 2012 Laboratory evolution of copper tolerant yeast strains. Microbial Cell Factories **11**: 1.

ADAMS, J., and F. ROSENZWEIG, 2014 Experimental microbial evolution: history and conceptual underpinnings. Genomics **104**: 393–398.

AKEN, B. L., P. ACHUTHAN, W. AKANNI, M. R. AMODE, F. BERNSDORFF, *et al.*, 2017 Ensembl 2017. Nucleic Acids Research **45**: D635–D642.

ALEXANDER, H. K., G. MARTIN, O. Y. MARTIN, and S. BONHOEFFER, 2014 Evolutionary rescue: linking theory for conservation and medicine. Evolutionary Applications **7**: 1161–1179.

ALEXANDER, R. P., G. FANG, J. ROZOWSKY, M. SNYDER, and M. B. GERSTEIN, 2010 Annotating non-coding regions of the genome. Nature Reviews. Genetics **11**: 559–571.

ALONSO, J. M., A. N. STEPANOVA, T. J. LEISSE, C. J. KIM, H. CHEN, *et al.*, 2003 Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301**: 653–657.

ANDERSON, J., J. FUNT, D. THOMPSON, S. PRABHU, A. SOCHA, *et al.*, 2010 Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. Current Biology **20**: 1383–1388.

ANDERSON, J. B., C. SIRJUSINGH, and N. RICKER, 2004 Haploidy, diploidy and evolution of antifungal drug resistance in *Saccharomyces cerevisiae*. Genetics **168**: 1915–23.

ANGERT, A. L., H. BRADSHAW JR, and D. W. SCHEMSKE, 2008 Using experimental evolution to investigate geographic range limits in monkeyflowers. Evolution **62**: 2660–2675.

ARNEGARD, M. E., M. D. MCGEE, B. MATTHEWS, K. B. MARCHINKO, G. L. CONTE, *et al.*, 2014 Genetics of ecological divergence during speciation. Nature **511**: 307–311.

ARTHINGTON-SKAGGS, B. A., H. JRADI, T. DESAI, and C. J. MORRISON, 1999 Quantitation of ergosterol content: novel method for determination of fluconazole susceptibility of *Candida albicans*. Journal of Clinical Microbiology **37**: 3332–3337.

AVERY, L., and S. WASSERMAN, 1992 Ordering gene function: the interpretation of epistasis in regulatory hierarchies. Trends in Genetics **8**: 312–316.

BABA, T., T. ARA, M. HASEGAWA, Y. TAKAI, Y. OKUMURA, *et al.*, 2006 Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology **2**: 2006.0008.

BAILEY, S. F., and T. BATAILLON, 2016 Can the experimental evolution programme help us elucidate the genetic basis of adaptation in nature? Molecular Ecology **25**: 203–218.

BALABAN, N. Q., J. MERRIN, R. CHAIT, L. KOWALIK, and S. LEIBLER, 2004 Bacterial persistence as a phenotypic switch. Science **305**: 1622–1625.

BALAKRISHNAN, R., J. PARK, K. KARRA, B. C. HITZ, G. BINKLEY, *et al.*, 2012 YeastMine–an integrated data warehouse for *Saccharomyces cerevisiae* data as a multipurpose tool-kit. Database **2012**: bar062.

BANDYOPADHYAY, S., M. MEHTA, D. KUO, M.-K. SUNG, R. CHUANG, *et al.*, 2010 Rewiring of genetic networks in response to DNA damage. Science **330**: 1385–1389.

BARKER, B., L. XU, and Z. GU, 2015 Dynamic epistasis under varying environmental perturbations. PloS One **10**: e0114911.

BARRICK, J. E., and R. E. LENSKI, 2013 Genome dynamics during experimental evolution. Nature Reviews. Genetics **14**: 827.

BARTON, N., and L. PARTRIDGE, 2000 Limits to natural selection. BioEssays **22**: 1075–1084.

BATES, D., M. MÄCHLER, B. BOLKER, and S. WALKER, 2015 Fitting linear mixed-effects models using lme4. Journal of Statistical Software **67**: 1–48.

BATESON, W., 1909 *Mendel's principles of heredity*. Cambridge University Press.

BEADLE, G. W., and E. L. TATUM, 1941 Genetic control of biochemical reactions in *Neurospora*. Proceedings of the National Academy of Sciences **27**: 499–506.

BELL, G., 2013 Evolutionary rescue and the limits of adaptation. Philosophical Transactions of the Royal Society B **368**: 20120080.

BELL, G., and S. COLLINS, 2008 Adaptation, extinction and global change. Evolutionary Applications **1**: 3–16.

BELL, G., and A. GONZALEZ, 2009 Evolutionary rescue can prevent extinction following environmental change. Ecology Letters **12**: 942–948.

BELLEN, H. J., R. W. LEVIS, Y. HE, J. W. CARLSON, M. EVANS-HOLM, *et al.*, 2011 The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. Genetics **188**: 731–743.

BEN-ARI, G., D. ZENVIRTH, A. SHERMAN, L. DAVID, M. KLUTSTEIN, *et al.*, 2006 Four linked genes participate in controlling sporulation efficiency in budding yeast. PLoS Genetics **2**: e195.

BERNARDES, J. P., R. B. STELKENS, and D. GREIG, 2017 Heterosis in hybrids within and between yeast species. Journal of Evolutionary Biology **30**: 538–548.

BIKARD, D., D. PATEL, C. LE METTÉ, V. GIORGI, C. CAMILLERI, *et al.*, 2009 Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. Science **323**: 623–626.

BLANQUART, F., G. ACHAZ, T. BATAILLON, and O. TENAILLON, 2014 Properties of selected mutations and genotypic landscapes under Fisher's geometric model. Evolution **68**: 3537–3554.

BLOWS, M. W., and A. A. HOFFMANN, 2005 A reassessment of genetic limits to evolutionary change. Ecology **86**: 1371–1384.

BOONE, C., H. BUSSEY, and B. J. ANDREWS, 2007 Exploring genetic interactions and networks with yeast. Nature Reviews. Genetics **8**: 437–449.

BORDENSTEIN, S. R., and M. D. DRAPEAU, 2001 Genotype-by-environment interaction and the Dobzhansky–Muller model of postzygotic isolation. Journal of Evolutionary Biology **14**: 490–501.

BOTSTEIN, D., and G. R. FINK, 2011 Yeast: An experimental organism for 21st century biology. Genetics **189**: 695–704.

BRESLOW, D. K., D. M. CAMERON, S. R. COLLINS, M. SCHULDINER, J. STEWART-ORNSTEIN, *et al.*, 2008 A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nature Methods **5**: 711–718.

BURCH, C. L., and L. CHAO, 1999 Evolution by small steps and rugged landscapes in the RNA virus phi6. Genetics **151**: 921–927.

BURKE, M. K., J. P. DUNHAM, P. SHAHRESTANI, K. R. THORNTON, M. R. ROSE, *et al.*, 2010 Genome-wide analysis of a long-term evolution experiment with *Drosophila*. Nature **467**: 587.

BUSKIRK, S. W., R. E. PEACE, and G. I. LANG, 2017 Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. Proceedings of the National Academy of Sciences **114**: 8330–8335.

BYRNE, A. B., M. T. WEIRAUCH, V. WONG, M. KOEVA, S. J. DIXON, *et al.*, 2007 A global analysis of genetic interactions in *Caenorhabditis elegans*. Journal of Biology **6**: 8.

C. ELEGANS DELETION MUTANT CONSORTIUM AND OTHERS, 2012 Large-scale screening for targeted knockouts in the *Caenorhabditis elegans* genome. G3: Genes, Genomes, Genetics **2**: 1415–1425.

CARLSON, S. M., C. J. CUNNINGHAM, and P. A. H. WESTLEY, 2014 Evolutionary rescue in a changing world. Trends in Ecology & Evolution **29**: 521–530.

CARRILLO-MUÑOZ, A. J., G. GIUSIANO, P. A. EZKURRA, and G. QUINDÓS, 2006 Antifungal agents: mode of action in yeast cells. Revista Española de Quimioterapia **19**: 130–139.

CASPETA, L., Y. CHEN, P. GHIACI, A. FEIZI, S. BUSKOV, *et al.*, 2014 Altered sterol composition renders yeast thermotolerant. Science **346**: 75–78.

CHATR-ARYAMONTRI, A., R. OUGHTRED, L. BOUCHER, J. RUST, C. CHANG, *et al.*, 2017 The BioGRID interaction database: 2017 update. Nucleic Acids Research **45**: D369–D379.

CHERRY, J. M., E. L. HONG, C. AMUNDSEN, R. BALAKRISHNAN, G. BINKLEY, *et al.*, 2011 *Saccharomyces* Genome Database: the genomics resource of budding yeast. Nucleic Acids Research **40**: D700–D705.

CHOU, H. H., J. BERTHET, and C. J. MARX, 2009 Fast growth increases the selective advantage of a mutation arising recurrently during evolution under metal limitation. PLoS Genetics **5**: e1000652.

CHOU, H.-H., H.-C. CHIU, N. F. DELANEY, D. SEGRÈ, and C. J. MARX, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science **332**: 1190–1192.

CHOU, H.-H., N. F. DELANEY, J. A. DRAGHI, and C. J. MARX, 2014 Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. PLoS Genetics **10**: e1004149–11.

CONRAD, T. M., A. R. JOYCE, M. K. APPLEBEE, C. L. BARRETT, B. XIE, *et al.*, 2009 Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biology **10**: R118.

CONTE, G. L., M. E. ARNEGARD, J. BEST, Y. F. CHAN, F. C. JONES, *et al.*, 2015 Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. Genetics **201**: 1189–1200.

CONTE, G. L., M. E. ARNEGARD, C. L. PEICHEL, and D. SCHLUTER, 2012 The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society B **279**: 5039–5047.

COSTANZO, M., B. VANDERSLUIS, E. N. KOCH, A. BARYSHNIKOVA, C. PONS, *et al.*, 2016 A global genetic interaction network maps a wiring diagram of cellular function. Science **353**: aaf1420–aaf1420.

COVO, S., C. M. PUCCIA, J. L. ARGUESO, D. A. GORDENIN, and M. A. RESNICK, 2014 The sister chromatid cohesion pathway suppresses multiple chromosome gain and chromosome amplification. Genetics **196**: 373–384.

COYNE, J., and H. ORR, 2004 *Speciation*. Sinauer Associates, Incorporated Publishers.

CREGG, J. M., 2007 DNA-mediated transformation. *Pichia* Protocols : 27–42.

CUTTER, A., 2005 Mutation and the experimental evolution of outcrossing in *Caenorhabditis elegans*. Journal of Evolutionary Biology **18**: 27–34.

DARWIN, C., 1859 *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life*. London: John Murray.

DARWIN, C., and A. WALLACE, 1858 On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. Zoological Journal of the Linnean Society **3**: 45–62.

DE VISSER, J. A. G. M., T. F. COOPER, and S. F. ELENA, 2011 The causes of epistasis. Proceedings of the Royal Society B **278**: 3617–3624.

DE VISSER, J. A. G. M., and D. E. ROZEN, 2005 Limits to adaptation in asexual populations. Journal of Evolutionary Biology **18**: 779–788.

DETTMAN, J. R., J. B. ANDERSON, and L. M. KOHN, 2008 Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus *Neurospora*. BMC Evolutionary Biology **8**: 35.

DETTMAN, J. R., C. SIRJUSINGH, L. M. KOHN, and J. B. ANDERSON, 2007 Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. Nature **447**: 585–588.

DEUTSCHBAUER, A. M., and R. W. DAVIS, 2005 Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nature Genetics **37**: 1333–1340.

DHAR, R., R. SÄGESSER, C. WEIKERT, J. YUAN, and A. WAGNER, 2011 Adaptation of *Saccharomyces cerevisiae* to saline stress through laboratory evolution. Journal of Evolutionary Biology **24**: 1135–1153.

DIXON, S. J., M. COSTANZO, A. BARYSHNIKOVA, B. ANDREWS, and C. BOONE, 2009 Systematic mapping of genetic interaction networks. Annual Review of Genetics **43**: 601–625.

DIXON, S. J., Y. FEDYSHYN, J. L. Y. KOH, T. S. K. PRASAD, C. CHAHWAN, *et al.*, 2008 Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. Proceedings of the National Academy of Sciences **105**: 16653–16658.

DOWELL, R. D., O. RYAN, A. JANSEN, D. CHEUNG, S. AGARWALA, *et al.*, 2010 Genotype to phenotype: a complex problem. Science **328**: 469–469.

EHRENREICH, I. M., J. BLOOM, N. TORABI, X. WANG, Y. JIA, *et al.*, 2012 Genetic architecture of highly complex chemical resistance traits across four yeast strains. PLoS Genetics **8**: e1002570.

ELROD, S. L., S. M. CHEN, K. SCHWARTZ, and E. O. SHUSTER, 2009 Optimizing sporulation conditions for different *Saccharomyces cerevisiae* strain backgrounds. Methods in Molecular Biology **557**: 21–26.

ENTIAN, K.-D., T. SCHUSTER, J. HEGEMANN, D. BECHER, H. FELDMANN, *et al.*, 1999 Functional analysis of 150 deletion mutants in *Saccharomyces cerevisiae* by a systematic approach. Molecular and General Genetics **262**: 683–702.

FERNANDES, A. R., M. PRIETO, and I. SÁ-CORREIA, 2000 Modification of plasma membrane lipid order and H+-ATPase activity as part of the response of *Saccharomyces cerevisiae* to cultivation under mild and high copper stress. Archives of Microbiology **173**: 262–268.

FERNANDES, A. R., and I. SÁ-CORREIA, 2001 The activity of plasma membrane H(+)-ATPase is strongly stimulated during *Saccharomyces cerevisiae* adaptation to growth under high copper stress, accompanying intracellular acidification. Yeast **18**: 511–521.

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

FOGEL, S., and J. W. WELCH, 1982 Tandem gene amplification mediates copper resistance in yeast. Proceedings of the National Academy of Sciences **79**: 5342–5346.

FOGEL, S., J. W. WELCH, G. CATHALA, and M. KARIN, 1983 Gene amplification in yeast: *CUP1* copy number regulates copper resistance. Current Genetics **7**: 347–355.

FOGLE, C. A., J. L. NAGLE, and M. M. DESAI, 2008 Clonal interference, multiple mutations and adaptation in large asexual populations. Genetics **180**: 2163–2173.

FRAÏSSE, C., P. A. GUNNARSSON, D. ROZE, N. BIERNE, and J. J. WELCH, 2016 The genetics of speciation: Insights from Fisher's geometric model. Evolution **70**: 1450–1464.

FROST, A., M. G. ELGORT, O. BRANDMAN, C. IVES, S. R. COLLINS, *et al.*, 2012 Functional repurposing revealed by comparing *S. pombe* and *S. cerevisiae* genetic interactions. Cell **149**: 1339–1352.

GARLAND, T., and M. R. ROSE, 2009 *Experimental evolution: concepts, methods, and applications of selection experiments*. University of California Press.

GERSTEIN, A. C., 2013 Mutational effects depend on ploidy level: all else is not equal. Biology Letters **9**: 20120614–20120614.

GERSTEIN, A. C., L. A. CLEATHERO, M. A. MANDEGAR, and S. P. OTTO, 2011 Haploids adapt faster than diploids across a range of environments. Journal of Evolutionary Biology **24**: 531–540.

GERSTEIN, A. C., A. KUZMIN, and S. P. OTTO, 2014 Loss-of-heterozygosity facilitates passage through Haldane's sieve for *Saccharomyces cerevisiae* undergoing adaptation. Nature Communications **5**: 3819.

GERSTEIN, A. C., D. S. LO, and S. P. OTTO, 2012 Parallel genetic changes and nonparallel gene-environment interactions characterize the evolution of drug resistance in yeast. Genetics **192**: 241–252.

GIAEVER, G., A. M. CHU, L. NI, C. CONNELLY, L. RILES, *et al.*, 2002 Functional profiling of the *Saccharomyces cerevisiae* genome. Nature **418**: 387–391.

GOMPEL, N., and B. PRUD'HOMME, 2009 The causes of repeated genetic evolution. Developmental Biology **332**: 36–47.

GOMPEL, N., B. PRUD'HOMME, P. J. WITTKOPP, V. A. KASSNER, and S. B. CARROLL, 2005 Chance caught on the wing: *cis*-regulatory evolution and the origin of pigment patterns in *Drosophila*. Nature **433**: 481–487.

GOMULKIEWICZ, R., and R. D. HOLT, 1995 When does evolution by natural selection prevent extinction? Evolution **49**: 201–207.

GOULD, S., 1989 *Wonderful Life: The Burgess Shale and the Nature of History*. Norton, New York.

GRADEN, J. A., and D. R. WINGE, 1997 Copper-mediated repression of the activation domain in the yeast Mac1p transcription factor. Proceedings of the National Academy of Sciences **94**: 5550–5555.

GRAMATES, L. S., S. J. MARYGOLD, G. D. SANTOS, J.-M. URBANO, G. ANTONAZZO, *et al.*, 2017 Flybase at 25: looking to the future. Nucleic Acids Research **45**: D663–D671.

GRESHAM, D., M. M. DESAI, C. M. TUCKER, H. T. JENQ, D. A. PAI, *et al.*, 2008 The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genetics **4**: e1000303.

GROSS, B. L., and L. H. RIESEBERG, 2005 The ecological genetics of homoploid hybrid speciation. Journal of Heredity **96**: 241–252.

HARMAND, N., R. GALLET, R. JABBOUR-ZAHAB, G. MARTIN, and T. LENORMAND, 2017 Fisher's geometrical model and the mutational patterns of antibiotic resistance across dose gradients. Evolution **71**: 23–37.

HARRISON, R., B. PAPP, C. PAL, S. G. OLIVER, and D. DELNERI, 2007 Plasticity of genetic interactions in metabolic networks of yeast. Proceedings of the National Academy of Sciences **104**: 2307–2312.

HARROW, J., A. FRANKISH, J. M. GONZALEZ, E. TAPANARI, M. DIEKHANS, *et al.*, 2012 Gencode: the reference human genome annotation for the encode project. Genome Research **22**: 1760–1774.

HARTMAN, J. L., B. GARVIK, and L. HARTWELL, 2001 Principles for the buffering of genetic variation. Science **291**: 1001–1004.

HERRON, M. D., and M. DOEBELI, 2013 Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. PLoS Biology **11**: e1001490.

HICKMAN, M. A., G. ZENG, A. FORCHE, M. P. HIRAKAWA, D. ABBEY, *et al.*, 2013 The 'obligate diploid' *Candida albicans* forms mating-competent haploids. Nature **494**: 55–59.

HO, C. H., L. MAGTANONG, S. L. BARKER, D. GRESHAM, S. NISHIMURA, *et al.*, 2009 A molecular barcoded yeast orf library enables mode-of-action analysis of bioactive compounds. Nature Biotechnology **27**: 369–377.

HOU, J., A. FRIEDRICH, J.-S. GOUNOT, and J. SCHACHERER, 2015 Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. Nature Communications **6**: 7214.

HUALA, E., A. W. DICKERMAN, M. GARCIA-HERNANDEZ, D. WEEMS, L. REISER, *et al.*, 2001 The *Arabidopsis* information resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Research **29**: 102–105.

HUSE, H. K., T. KWON, J. E. A. ZLOSNIK, D. P. SPEERT, E. M. MARCOTTE, *et al.*, 2010 Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations *in vivo*. mBio **1**: e00199–10.

HVORECNY, K. L., and G. PRELICH, 2010 A systematic cen library of the *Saccharomyces cerevisiae* genome. Yeast **27**: 861–865.

JASNOS, L., K. TOMALA, D. PACZESNIAK, and R. KORONA, 2008 Interactions between stressful environment and gene deletions alleviate the expected average loss of fitness in yeast. Genetics **178**: 2105–2111.

JAYAKODY, L. N., N. HAYASHI, and H. KITAGAKI, 2011 Identification of glycolaldehyde as the key inhibitor of bioethanol fermentation by yeast and genome-wide analysis of its toxicity. Biotechnology Letters **33**: 285–292.

JELIER, R., J. I. SEMPLE, R. GARCIA-VERDUGO, and B. LEHNER, 2011 Predicting phenotypic variation in yeast from individual genome sequences. Nature Genetics **43**: 1270–1274.

JENSEN-PERGAKES, K. L., M. A. KENNEDY, N. D. LEES, R. BARBUCH, C. KOEGEL, *et al.*, 1998 Sequencing, disruption, and characterization of the *Candida albicans* sterol methyltransferase (*ERG6*) gene: drug susceptibility studies in *erg6* mutants. Antimicrobial Agents and Chemotherapy **42**: 1160–1167.

JONES, G. M., J. STALKER, S. HUMPHRAY, A. WEST, T. COX, *et al.*, 2008 A systematic library for comprehensive overexpression screens in *Saccharomyces cerevisiae*. Nature Methods **5**: 239–241.

JOSHI, C. J., and A. PRASAD, 2014 Epistatic interactions among metabolic genes depend upon environmental conditions. Molecular Biosystems **10**: 2578–2589.

KACHROO, A. H., J. M. LAURENT, A. AKHMETOV, M. SZILAGYI-JONES, C. D. MCWHITE, *et al.*, 2017 Systematic bacterialization of yeast genes identifies a near-universally swappable pathway. eLife **6**: e25093.

KAO, K. C., and G. SHERLOCK, 2008 Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. Nature Genetics **40**: 1499–1504.

KASSEN, R., 2014 *Experimental evolution and the nature of biodiversity*. Roberts.

KAWECKI, T. J., R. E. LENSKI, D. EBERT, B. HOLLIS, I. OLIVIERI, *et al.*, 2012 Experimental evolution. Trends in Ecology & Evolution **27**: 547–560.

KESELER, I. M., A. MACKIE, A. SANTOS-ZAVALETA, R. BILLINGTON, C. BONAVIDES-MARTÍNEZ, *et al.*, 2017 The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Research **45**: D543–D550.

KHAN, A. I., D. M. DINH, D. SCHNEIDER, R. E. LENSKI, and T. F. COOPER, 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. Science **332**: 1193–1196.

KINSELLA, R. J., A. KÄHÄRI, S. HAIDER, J. ZAMORA, G. PROCTOR, *et al.*, 2011 Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database **2011**: bar030–bar030.

KISHONY, R., and S. LEIBLER, 2003 Environmental stresses can alleviate the average deleterious effect of mutations. Journal of Biology **2**: 14.

KOMSTA, L., 2011 *Outliers: Tests for outliers*. R package version 0.14.

KONO, K., A. AL-ZAIN, L. SCHROEDER, M. NAKANISHI, and A. E. IKUI, 2016 Plasma membrane/cell wall perturbation activates a novel cell cycle checkpoint during G1 in *Saccharomyces cerevisiae*. Proceedings of the National Academy of Sciences **113**: 6910–6915.

KRYAZHIMSKIY, S., D. RICE, E. JERISON, and M. DESAI, 2014 Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science **344**: 1519–1522.

KUMAR, A., N. BELOGLAZOVA, C. BUNDALOVIC-TORMA, S. PHANSE, V. DEINEKO, *et al.*, 2016 Conditional epistatic interaction maps reveal global functional rewiring of genome integrity pathways in *Escherichia coli*. Cell Reports **14**: 648–661.

KVITEK, D. J., and G. SHERLOCK, 2011 Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genetics **7**: e1002056.

LACHAPELLE, J., and G. BELL, 2012 Evolutionary rescue of sexual and asexual populations in a deteriorating environment. Evolution **66**: 3508–3518.

LANDRY, C. R., J. P. TOWNSEND, D. L. HARTL, and D. CAVALIERI, 2006 Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. Molecular Ecology **15**: 575–591.

LANG, G. I., D. P. RICE, M. J. HICKMAN, E. SODERGREN, G. M. WEINSTOCK, *et al.*, 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature **500**: 571–574.

LARKIN, M. A., G. BLACKSHIELDS, N. BROWN, R. CHENNA, P. A. MCGETTIGAN, *et al.*, 2007 Clustal W and Clustal X version 2.0. Bioinformatics **23**: 2947–2948.

LEE, A. Y., R. P. ST ONGE, M. J. PROCTOR, I. M. WALLACE, A. H. NILE, *et al.*, 2014 Mapping the cellular response to small molecules using chemogenomic fitness signatures. Science **344**: 208–211.

LEE, I., Z. LI, and E. M. MARCOTTE, 2007 An improved, bias-reduced probabilistic functional gene network of baker's yeast, *Saccharomyces cerevisiae*. PloS One **2**: e988.

LEES, N. D., B. SKAGGS, D. R. KIRSCH, and M. BARD, 1995 Cloning of the late genes in the ergosterol biosynthetic pathway of *Saccharomyces cerevisiae* – a review. Lipids **30**: 221–226.

LEHNER, B., C. CROMBIE, J. TISCHLER, A. FORTUNATO, and A. G. FRASER, 2006 Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nature Genetics **38**: 896–903.

LENSKI, R. E., 2017 Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. The ISME Journal **11**: 2181–2194.

LENTH, R. V., 2016 Least-squares means: The R package lsmeans. Journal of Statistical Software **69**: 1–33.

LEXER, C., Z. LAI, and L. H. RIESEBERG, 2004 Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. New Phytologist **161**: 225–233.

LI, H., B. HANDSAKER, A. WYSOKER, T. FENNELL, J. RUAN, *et al.*, 2009 The sequence alignment/map format and SAMtools. Bioinformatics **25**: 2078–2079.

LINDSEY, H. A., J. GALLIE, S. TAYLOR, and B. KERR, 2013 Evolutionary rescue from extinction is contingent on a lower rate of environmental change. Nature **494**: 463–467.

LITI, G., 2015 The natural history of model organisms: The fascinating and secret wild life of the budding yeast *S. cerevisiae*. Elife **4**: e05835.

LOBKOVSKY, A. E., and E. V. KOONIN, 2012 Replaying the tape of life: quantification of the predictability of evolution. Frontiers in Genetics **3**: 246.

LONG, A., G. LITI, A. LUPTAK, and O. TENAILLON, 2015 Elucidating the molecular architecture of adaptation via evolve and resequence experiments. Nature Reviews. Genetics **16**: 567.

LOSOS, J. B., 1992 The evolution of convergent structure in Caribbean *Anolis* communities. Systems Biology **41**: 403–420.

LOURENÇO, M., R. S. RAMIRO, D. GÜLERESI, J. BARROSO-BATISTA, K. B. XAVIER, *et al.*, 2016 A mutational hotspot and strong selection contribute to the order of mutations selected for during *Escherichia coli* adaptation to the gut. PLoS Genetics **12**: e1006420.

LYNCH, M., W. SUNG, K. MORRIS, N. COFFEY, and C. LANDRY, 2008 A genome-wide view of the spectrum of spontaneous mutations in yeast. Proceedings of the National Academy of Sciences **105**: 9272–9277.

MABLE, B., 2001 Ploidy evolution in the yeast *Saccharomyces cerevisiae*: a test of the nutrient limitation hypothesis. Journal of Evolutionary Biology **14**: 157–170.

MACLEAN, R. C., A. R. HALL, G. G. PERRON, and A. BUCKLING, 2010 The population genetics of antibiotic resistance: integrating molecular mechanisms and treatment contexts. Nature Reviews. Genetics **11**: 405–414.

MAHESHWARI, S., and D. A. BARBASH, 2011 The genetics of hybrid incompatibilities. Annual Review of Genetics **45**: 331–355.

MALLET, J., 2007 Hybrid speciation. Nature **446**: 279–283.

MANCEAU, M., V. S. DOMINGUES, C. R. LINNEN, E. B. ROSENBLUM, and H. E. HOEKSTRA, 2010 Convergence in pigmentation at multiple levels: mutations, genes and function. Philosophical Transactions of the Royal Society B **365**: 2439–2450.

MANDEGAR, M., and S. OTTO, 2007 Mitotic recombination counteracts the benefits of genetic segregation. Proceedings of the Royal Society B **274**: 1301–1307.

MARTIN, A., and V. ORGOGOZO, 2013 The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. Evolution **67**: 1235–1250.

MATSUO, T., S. SUGAYA, J. YASUKAWA, T. AIGAKI, and Y. FUYAMA, 2007 Odorant-binding proteins obp57d and obp57e affect taste perception and host-plant preference in *Drosophila sechellia*. PLoS Biology **5**: e118.

MAURICIO, R., 1998 Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. The American Naturalist **151**: 20–28.

MCCOURT, P., H.-Y. LIU, J. E. PARKER, C. GALLO-EBERT, M. DONIGAN, *et al.*, 2016 Proper sterol distribution is required for *Candida albicans* hyphal formation and virulence. G3: Genes, Genomes, Genetics **6**: 3455–3465.

MCDONALD, M. J., D. P. RICE, and M. M. DESAI, 2016 Sex speeds adaptation by altering the dynamics of molecular evolution. Nature **531**: 233–236.

MCDOWALL, M. D., M. A. HARRIS, A. LOCK, K. RUTHERFORD, D. M. STAINES, *et al.*, 2014 Pombase 2015: updates to the fission yeast database. Nucleic Acids Research **43**: D656–D661.

MERZ, S., and B. WESTERMANN, 2009 Genome-wide deletion mutant analysis reveals genes required for respiratory growth, mitochondrial genome maintenance and mitochondrial protein synthesis in *Saccharomyces cerevisiae*. Genome Biology **10**: R95.

MEYER, J. R., D. T. DOBIAS, J. S. WEITZ, J. E. BARRICK, R. T. QUICK, *et al.*, 2012 Repeatability and contingency in the evolution of a key innovation in phage lambda. Science **335**: 428–432.

MILLER, C., P. JOYCE, and H. WICHMAN, 2011 Mutational effects and population dynamics during viral adaptation challenge current models. Genetics **187**: 185.

MIRA, P. M., J. C. MEZA, A. NANDIPATI, and M. BARLOW, 2015 Adaptive landscapes of resistance genes change as antibiotic concentrations change. Molecular Biology and Evolution **32**: 2707–2715.

MORTIMER, R. K., and J. R. JOHNSTON, 1986 Genealogy of principal strains of the yeast genetic stock center. Genetics **113**: 35–43.

MUKHOPADHYAY, K., A. KOHLI, and R. PRASAD, 2002 Drug susceptibilities of yeast cells are affected by membrane lipid composition. Antimicrobial Agents and Chemotherapy **46**: 3695–3705.

MUSSO, G., M. COSTANZO, M. HUANGFU, A. M. SMITH, J. PAW, *et al.*, 2008 The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. Genome Research **18**: 1092–1099.

NOSIL, P., and D. SCHLUTER, 2011 The genes underlying the process of speciation. Trends in Ecology & Evolution **26**: 160–167.

ONO, J., A. C. GERSTEIN, and S. P. OTTO, 2016 Data from: Widespread genetic incompatibilities between first-step mutations during parallel adaptation of *Saccharomyces cerevisiae* to a common environment. Dryad Digital Repository. Openly available via http://dx.doi.org/10.5061/dryad.vs370.

ONO, J., A. C. GERSTEIN, and S. P. OTTO, 2017 Widespread genetic incompatibilities between first-step mutations during parallel adaptation of *Saccharomyces cerevisiae* to a common environment. PLoS Biology **15**: e1002591.

ORR, H. A., 1995 The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics **139**: 1805–1813.

ORR, H. A., and A. J. BETANCOURT, 2001 Haldane's sieve and adaptation from the standing genetic variation. Genetics **157**: 875–884.

ORR, H. A., and S. P. OTTO, 1994 Does diploidy increase the rate of adaptation? Genetics **136**: 1475–1480.

ORR, H. A., and R. L. UNCKLESS, 2008 Population extinction and the genetics of adaptation. The American Naturalist **172**: 160–169.

OTTO, S. P., and J. WHITTON, 2000 Polyploid incidence and evolution. Annual Review of Genetics **34**: 401–437.

PAPP, B., R. A. NOTEBAART, and C. PAL, 2011 Systems-biology approaches for predicting genomic evolution. Nature Reviews. Genetics **12**: 591–602.

PAYEN, C., A. B. SUNSHINE, G. T. ONG, J. L. POGACHAR, W. ZHAO, *et al.*, 2016 High-throughput identification of adaptive mutations in experimentally evolved yeast populations. PLoS Genetics **12**: e1006339.

PEÑA, M., J. LEE, and D. THIELE, 1999 A delicate balance: homeostatic control of copper uptake and distribution. Journal of Nutrition **129**: 1251–1260.

POELWIJK, F. J., D. J. KIVIET, D. M. WEINREICH, and S. J. TANS, 2007 Empirical fitness landscapes reveal accessible evolutionary paths. Nature **445**: 383.

POELWIJK, F. J., S. TĂNASE-NICOLA, D. J. KIVIET, and S. J. TANS, 2011 Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. Journal of Theoretical Biology **272**: 141–144.

PRÁ, D., S. I. R. FRANKE, R. GIULIAN, M. L. YONEAMA, J. F. DIAS, *et al.*, 2008 Genotoxicity and mutagenicity of iron and copper in mice. Biometals **21**: 289–297.

PRESGRAVES, D. C., 2010 The molecular evolutionary basis of species formation. Nature Reviews. Genetics **11**: 175.

PRUD'HOMME, B., N. GOMPEL, A. ROKAS, V. A. KASSNER, T. M. WILLIAMS, *et al.*, 2006 Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature **440**: 1050.

R CORE TEAM, 2015 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

R CORE TEAM, 2016 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

REMOLD, S. K., and R. E. LENSKI, 2004 Pervasive joint influence of epistasis and plasticity on mutational effects in *Escherichia coli*. Nature Genetics **36**: 423–426.

REPLANSKY, T., V. KOUFOPANOU, D. GREIG, and G. BELL, 2008 *Saccharomyces sensu stricto* as a model system for evolution and ecology. Trends in Ecology & Evolution **23**: 494–501.

ROBINSON, M. D., J. GRIGULL, N. MOHAMMAD, and T. R. HUGHES, 2002 FunSpec: a web-based cluster interpreter for yeast. BMC Bioinformatics **3**: 35.

ROGUEV, A., S. BANDYOPADHYAY, M. ZOFALL, K. ZHANG, T. FISCHER, *et al.*, 2008 Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science **322**: 405–410.

ROTH, F. P., H. D. LIPSHITZ, and B. J. ANDREWS, 2009 Q&A: Epistasis. Journal of Biology **8**: 35.

ROZEN, D., J. DE VISSER, and P. GERRISH, 2002 Fitness effects of fixed beneficial mutations in microbial populations. Current Biology **12**: 1040–1045.

RYAN, C. J., P. CIMERMANČIČ, Z. A. SZPIECH, A. SALI, R. D. HERNANDEZ, *et al.*, 2013 High-resolution network biology: connecting sequence with function. Nature Reviews. Genetics **14**: 865–879.

RYAN, C. J., A. ROGUEV, K. PATRICK, J. XU, H. JAHARI, *et al.*, 2012 Hierarchical modularity and the evolution of genetic interactomes across species. Molecular Cell **46**: 691–704.

SAMBROOK, J., and D. W. RUSSELL, 2001 *Molecular Cloning: A laboratory manual,. 3rd ed*. Cold Spring Harbor Laboratory Press, New York, USA.

SANGLARD, D., F. ISCHER, T. PARKINSON, D. FALCONER, and J. BILLE, 2003 *Candida albicans* mutations in the ergosterol biosynthetic pathway and resistance to several antifungal agents. Antimicrobial Agents and Chemotherapy **47**: 2404–2412.

SCHENK, M., I. SZENDRO, M. SALVERDA, J. KING, and J. DE VISSER, 2013 Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. Molecular Biology and Evolution **30**: 1779–1787.

SCHLUTER, D., 2009 Evidence for ecological speciation and its alternative. Science **323**: 737–741.

SCHLUTER, D., E. A. CLIFFORD, M. NEMETHY, and J. S. MCKINNON, 2004 Parallel evolution and inheritance of quantitative traits. The American Naturalist **163**: 809–822.

SCHNEEBERGER, K., 2014 Using next-generation sequencing to isolate mutant genes from forward genetic screens. Nature Reviews. Genetics **15**: 662–676.

SCHOUSTRA, S., S. HWANG, J. KRUG, and J. A. G. M. DE VISSER, 2016 Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. Proceedings of the Royal Society B **283**: 20161376.

SELMECKI, A. M., K. DULMAGE, L. E. COWEN, J. B. ANDERSON, and J. BERMAN, 2009 Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. PLoS Genetics **5**: e1000705.

SHERMAN, F., 2002 Getting started with yeast. Methods in Enzymology **350**: 3–41.

SIONOV, E., H. LEE, Y. C. CHANG, and K. J. KWON-CHUNG, 2010 *Cryptococcus neoformans* overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. PLoS Pathogens **6**: e1000848.

SMITH, J. M., R. BURIAN, S. KAUFFMAN, P. ALBERCH, J. CAMPBELL, *et al.*, 1985 Developmental constraints and evolution: A perspective from the mountain lake conference on development and evolution. The Quarterly Review of Biology **60**: 265–287.

SOARES, E. V., 2011 Flocculation in *Saccharomyces cerevisiae*: a review. Journal of Applied Microbiology **110**: 1–18.

ST ONGE, R. P., R. MANI, J. OH, M. PROCTOR, E. FUNG, *et al.*, 2007 Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nature Genetics **39**: 199–206.

STEIN, L., P. STERNBERG, R. DURBIN, J. THIERRY-MIEG, and J. SPIETH, 2001 Wormbase: network access to the genome and biology of *Caenorhabditis elegans*. Nucleic Acids Research **29**: 82–86.

STELKENS, R. B., M. A. BROCKHURST, G. D. D. HURST, and D. GREIG, 2014 Hybridization facilitates evolutionary rescue. Evolutionary Applications **7**: 1209–1217.

STERN, D. L., 2000 Perspective: evolutionary developmental biology and the problem of variation. Evolution **54**: 1079–1091.

STERN, D. L., 2013 The genetic causes of convergent evolution. Nature Reviews. Genetics **14**: 751–764.

STROOBANTS, A., J.-M. DELROISSE, F. DELVIGNE, J. DELVA, D. PORTETELLE, *et al.*, 2008 Isolation and biomass production of a *Saccharomyces cerevisiae* strain binding copper and zinc ions. Applied Biochemistry and Biotechnology **157**: 85–97.

SZATHMÁRY, E., 1993 Do deleterious mutations act synergistically? Metabolic control theory provides a partial answer. Genetics **133**: 127–132.

TENAILLON, O., A. RODRÍGUEZ-VERDUGO, R. L. GAUT, P. MCDONALD, A. F. BENNETT, *et al.*, 2012 The molecular diversity of adaptive convergence. Science **335**: 457–461.

TESTE, M.-A., M. DUQUENNE, J. M. FRANCOIS, and J.-L. PARROU, 2009 Validation of reference genes for quantitative expression analysis by real-time RT-PCR in *Saccharomyces cerevisiae*. BMC Molecular Biology **10**: 99.

THOMPSON, O., M. EDGLEY, P. STRASBOURGER, S. FLIBOTTE, B. EWING, *et al.*, 2013 The million mutation project: a new approach to genetics in *Caenorhabditis elegans*. Genome Research **23**: 1749–1762.

TISCHLER, J., B. LEHNER, and A. G. FRASER, 2008 Evolutionary plasticity of genetic interaction networks. Nature Genetics **40**: 390–391.

TKESHELASHVILI, L., T. MCBRIDE, K. SPENCE, and L. LOEB, 1991 Mutation spectrum of copper-induced DNA damage. Journal of Biological Chemistry **266**: 6401–6406.

TONG, Z., M.-S. KIM, A. PANDEY, and P. J. ESPENSHADE, 2014 Identification of candidate substrates for the Golgi Tul1 E3 ligase using quantitative diGly proteomics in yeast. Molecular and Cellular Proteomics **13**: 2871–2882.

WANG, A. D., N. P. SHARP, C. C. SPENCER, K. TEDMAN-AUCOIN, and A. F. AGRAWAL, 2009 Selection, epistasis, and parent-of-origin effects on deleterious mutations across environments in *Drosophila melanogaster*. The American Naturalist **174**: 863–874.

WEINREICH, D., R. WATSON, and L. CHAO, 2005 Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution **59**: 1165–1174.

WEIRAUCH, M. T., A. YANG, M. ALBU, A. G. COTE, A. MONTENEGRO-MONTERO, *et al.*, 2014 Determination and inference of eukaryotic transcription factor sequence specificity. Cell **158**: 1431–1443.

WELCH, J., D. MALONEY, and S. FOGEL, 1991 Gene conversions within the *Cup1*^{r} region from heterologous crosses in *Saccharomyces cerevisiae*. Molecular and General Genetics **229**: 261–266.

WENGER, J. W., J. PIOTROWSKI, S. NAGARAJAN, K. CHIOTTI, G. SHERLOCK, *et al.*, 2011 Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. PLoS Genetics **7**: e1002202.

WHITLOCK, M. C., P. C. PHILLIPS, F. B.-G. MOORE, and S. J. TONSOR, 1995 Multiple fitness peaks and epistasis. Annual Review of Ecology and Systematics **26**: 601–629.

WONG, A., N. RODRIGUE, and R. KASSEN, 2012 Genomics of adaptation during experimental evolution of the opportunistic pathogen *Pseudomonas aeruginosa*. PLoS Genetics **8**: e1002928.

WOODS, R., 1971 Nystatin-resistant mutants of yeast: alterations in sterol content. Journal of Bacteriology **108**: 69–73.

YAMPOLSKY, L. Y., and A. STOLTZFUS, 2005 The exchangeability of amino acids in proteins. Genetics **170**: 1459–1472.

YONA, A. H., Y. S. MANOR, R. H. HERBST, G. H. ROMANO, A. MITCHELL, *et al.*, 2012 Chromosomal duplication is a transient evolutionary solution to stress. Proceedings of the National Academy of Sciences **109**: 21010–21015.

ZEYL, C., T. VANDERFORD, and M. CARTER, 2003 An evolutionary advantage of haploidy in large yeast populations. Science **299**: 555–8.

ZHANG, H., A. F. ZEIDLER, W. SONG, C. M. PUCCIA, E. MALC, *et al.*, 2013 Gene copy-number variation in haploid and diploid strains of the yeast *Saccharomyces cerevisiae*. Genetics **193**: 785–801.

ZHAO, Y., P. K. STROPE, S. G. KOZMIN, J. H. MCCUSKER, F. S. DIETRICH, *et al.*, 2014 Structures of naturally evolved *CUP1* tandem arrays in yeast indicate that these arrays are generated by unequal nonhomologous recombination. G3: Genes, Genomes, Genetics **4**: 2259–2269.

ZHU, Y. O., G. SHERLOCK, and D. A. PETROV, 2016 Whole genome analysis of 132 clinical *Saccharomyces cerevisiae* strains reveals extensive ploidy variation. G3: Genes, Genomes, Genetics **6**: 2421–2434.

ZINSER, E., F. PALTAUF, and G. DAUM, 1993 Sterol composition of yeast organelle membranes and subcellular distribution of enzymes involved in sterol metabolism. Journal of Bacteriology **175**: 2853–2858.

Appendix

Appendix A

Appendix for Chapter 2: Too much of a good thing: The unique and repeated paths toward copper adaptation

A.1 Supplementary Methods

A.1.1 Quantitative real-time PCR (qPCR)

To test whether levels of *CUP1* inferred from *in silico* qPCR were consistent with levels of *CUP1* transcription, we assayed RNA levels using quantitative real-time PCR (qPCR). We chose 10 CBM lines that spanned the range of *CUP1* copy number (from lowest to highest): CBM16, CBM22, CBM24, CBM37, CBM2, CBM14, CBM51, CBM4, CBM34, and CBM13. For each line and BY4741, culture was struck from frozen onto a YPD plate and grown at 30°C for 48h. A single colony of each CBM line and two colonies of BY4741 were inoculated into 1mL YPD + 5.5mM copper (a lower concentration was used to allow growth of all lines, including BY4741) and grown for 12 hours at 30°C with shaking, at which point RNA was isolated using the RNEasy Mini Kit from Qiagen, following the yeast protocol. cDNA was reverse transcribed from 500ng of RNA using MultiScribe reverse transcriptase (Life Technologies) and oligo d(T) primers.

Oligonucleotides for qPCR (Table 2.2) were designed using Primer Express (ABI). mRNA levels of *TAF10* were used for normalization because *TAF10* has stable expression across strains and conditions (TESTE *et al.* 2009). cDNA was diluted 100-fold for *CUP1*, but not *TAF10*, to account for differences in their abundance in the samples. All qPCR reactions were performed using an ABI7000 sequence detection system (Applied Biosystems, Inc.). The reaction volume was 22μL containing 10pmol of each primer and 2μL of each sample with 10μL of 2X SYBR Green Master Mix (Applied Biosystems Inc.). Reaction conditions were 1 cycle of 50°C for 2 min., 1 cycle of 94°C for 10 min., and 40 cycles of 95°C for 15 sec., 60°C for 1 min.

To generate standard curves against which cDNA concentrations could be measured, we obtained cDNA from BY4741, which was then diluted five times, each time using a five-fold dilution, followed by qPCR using the primers for *TAF10* and *CUP1*. Standard curves for BY4741 were plotted such that a 1:1 relationship

between fluorescence and sample concentration yields an expected slope of $log_2(10) = -3.32$. All standard curves fit the data well (r^2 >0.99), with slopes between -3.1 and -3.3. *TAF10* and *CUP1* expression levels for every other strain were then measured against their respective standard curves, and then *CUP1* levels were divided by *TAF10* levels to control for variation among samples in total cDNA concentrations. All qPCR experiments were performed with two technical replicates.

A.1.2 Tetrad analyses

To separate the effects of single mutations from other mutations present in the evolved lines (including extra copies of *CUP1*), we crossed all of the CBM lines with BY4739 (*MAT*α *leu2*$\triangle$*0 lys2*$\triangle$*0 ura3*$\triangle$*0*), which has a common genotype yet opposite mating type and different auxotrophies than BY4741, the progenitor of our lines. Cells of both mating types were allowed time to mate overnight on a YPD plate before being struck onto plates lacking histidine and lysine, selecting for diploids. Single colonies were then struck onto selection plates a second time to ensure they were diploids. Culture was taken directly from these plates and frozen in 15% glycerol.

To isolate single mutations, we attempted to sporulate the CBM×BY4739 lines that contained each common mutation or aneuploidy and the fewest number of additional mutations ($\sim$1/3 of the lines). We encountered substantial difficulties in obtaining tetrads from our strains; BY4741, a derivative of S288c, is known to be a poor sporulator (BEN-ARI *et al.* 2006; DEUTSCHBAUER and DAVIS 2005). With a subset of the lines, we attempted to maximize sporulation rates using a variety of different protocols including all combinations of YPD, 1% YPA or 6% YEPD for pre-sporulation, liquid or plates, and PSP2, 1% KAc, or CSHSPO as the sporulation medium, all in liquid (see ELROD *et al.* 2009 for media details). Most combinations were tried at 30°C, but the CSHSPO combinations were also attempted at 25°C and 37°C. In all cases, frozen culture was struck onto YPD plates and grown for 48 hours at 30°C to isolate a single colony for sporulation. Ultimately, we obtained the most success using YPD liquid as a pre-sporulation medium followed by washing 100μL of overnight culture with dH_2O and then plating on 1% KAc at 20°C or 25°C for up to 30 days. Unfortunately, however, we remained unable to sporulate the majority of lines. In particular, despite many attempts, no tetrads were obtained for CBM16 (*PMA1* mutation plus chrII aneuploidy), CBM26 (*PMA1* mutation plus chrI, chrV and chrVIII aneuploidy), CBM29 (*PMA1* mutation plus chrII aneuploidy), CBM47 (*VTC1* mutation), or CBM55 (no mutation identified other than extra copies of *CUP1*).

We were able to sporule CBM2 (chrII aneuploidy), CBM14 (*MAM3* mutation), CBM25 (*MLP1* and *ENA5* mutations), and CBM34 (*VTC4* mutation). CBM25 was not initially chosen for tetrad dissection but was dissected as a contaminate of CBM22 (*VTC1* plus chrVIII and chrXVI aneuploidy), as detected by subsequent sequencing. CBM25 contaminating cells were likely positively selected during the sporulation procedure given that the aneuploid lines in our experiment, like CBM22, had very low sporulation rates. The resulting tetrads were dissected by micromanipulation on YPD plates. The spores were allowed to germinate and grow at 30°C for 3 days before each dissection plate was replica plated to test for mating types and auxotrophies. All tetrads were verified for 2:2 segregation of auxotrophies (except the aneuploid CBM2 - see below) and mating type. Once confirmed, the colonies obtained from each spore were frozen in 15%

glycerol. We numbered the dissected tetrads (t1 up to t11) and lettered each haploid spore (a-d).

The genotype of resulting spores was then determined. For CBM14 tetrad lines, *MAM3* was amplified by PCR (primers in Table 2.2), and the product was digested with EcoRV (Fermentas), which specifically cuts the mutant allele at GATATC. The results (one band versus two) were visualized on a 2% agarose gel. CBM25 spores were sequenced on Illumina HiSeq 2000, which is when the strain was discovered to be CBM25 (bearing a mutation in *MLP1* and *ENA5*), not CBM22. For CBM34 spores, *VTC4* was PCR amplified (Table 2.2) from genomic DNA and the fragment was Sanger sequenced using the forward primer and aligned to the reference sequence using ClustalW at EMBL-EBI (LARKIN *et al.* 2007). All SNPs showed the expected 2:2 segregation pattern.

The segregation pattern for the additional copy of chrII in CBM2 spores was determined for three of the tetrads (t1, t2 and t5) by the presence of the *LYS2* alleles, as detected by PCR. The *LYS2* gene is located on chrII, and the mated diploid from CBM2 carried two functioning copies of the gene (from CBM2) and one copy of the lys2$\triangle$0 allele (from BY4739). Primers were designed to flank the *LYS2* gene. The forward primer was designed 538bp upstream of the start site and the reverse primer was designed 487bp downstream of the stop codon in order to easily detect the deletion by band size (full gene = 5199bp, deletion allele = 708bp) (primers in Table 2.2). For t1 and t2, the two functional copies were inferred to be in the same cell due to the 2:2 segregation of the deletion and wild type alleles. For t5, one functional allele had segregated to each cell and the two aneuploid cells were determined based on PCR detection of the presence of the deletion. Corresponding phenotypes were verified by plating on medium lacking lysine. The segregation pattern for the additional copy of chrII in t3 was determined by Illumina sequencing, followed by calculating the total depth of coverage for each chromosome, as described above.

Southern blots with *CUP1* specific probes were performed to quantify the segregation patterns of *CUP1* among the spores. DNA concentration of genomic DNA isolated from each analyzed spore was measured in triplicate with the Qubit fluorometer (Invitrogen). Based on the average concentration, $2\mu g$ of each sample was loaded into a 1% agarose gel and run at 120V. DNA in the gel was denatured in a NaOH buffer and transferred to a nylon membrane (Hybond N+, GE Healthcare) using capillary transfer in 20x SSC buffer, affixing the DNA to the membrane by baking at 80°C for 2 hours. The membrane was incubated overnight at 57°C, with two biotin labelled probes (Table 2.2). The membrane was then washed in 2x SSC + 0.1% SDS buffer at 56°C 3 times for 15min. Probe binding was visualized using the North2South chemiluminescent detection kit (Thermo Scientific). Blots were exposed onto CL-XPosure Film (Thermo Scientific) for 30sec. to 1min. and developed in a Kodak X-ray film processor. We isolated genomic DNA and ran a Southern blot on three separate occasions for each spore. Controls (BY4741, BY4739, original CBM line) were always run in duplicate on the same gel as related spores. Band intensity was quantified in ImageJ (ABRAMOFF *et al.* 2004) using the "background corrected density" macro (`http://rsb.info.nih.gov/ij/macros/BackgroundCorrectedDensity.txt`).

A.2 Supporting Tables

Table A.1: Date of isolation for CBM lines. 56 putative mutation lines were isolated from three deep-well boxes (A-C) following exposure to copper (started on 19 Jan 2011). Culture from each well showing growth was streaked onto a YPD plate and assessed for colony size. Eight colonies were randomly chosen from all lines that grew normally on the YPD plates and assayed for growth in copper12. A single copper-resistant colony was chosen from each of the 34 remaining putative mutation lines. All lines were subsequently streaked onto YPG plates to assay respiratory capacity.

Date of isolation	CBM Line	Box	copper12 growth (# of 8)	YPG growth
27 Jan 2011	CBM1	A	7	✓
27 Jan 2011	CBM2	B	6	✓
27 Jan 2011	CBM3	B	3	✓
27 Jan 2011	CBM4	B	7	✓
27 Jan 2011	CBM5	C	5	✓
28 Jan 2011	CBM6	A	4	✓
28 Jan 2011	CBM7	A	6	✓
28 Jan 2011	CBM8	A	0†	petite
28 Jan 2011	CBM9	B	-‡	✓
28 Jan 2011	CBM10	B	-	petite
28 Jan 2011	CBM11	B	8	✓
28 Jan 2011	CBM12	B	-	petite
28 Jan 2011	CBM13	B	8	✓
28 Jan 2011	CBM14	C	7	✓
28 Jan 2011	CBM15	C	-	petite
28 Jan 2011	CBM16	C	7*	petite
28 Jan 2011	CBM17	C	6	✓
28 Jan 2011	CBM18	C	7	✓
29 Jan 2011	CBM19	A	-	petite
29 Jan 2011	CBM20	B	8*	petite
29 Jan 2011	CBM21	B	7	✓
29 Jan 2011	CBM22	C	6	✓
29 Jan 2011	CBM23	C	no growth on YPD	-
29 Jan 2011	CBM24	C	8	✓
30 Jan 2011	CBM25	A	4	✓
30 Jan 2011	CBM26	A	6	✓
30 Jan 2011	CBM27	A	-‡	✓
30 Jan 2011	CBM28	A	-‡	✓
30 Jan 2011	CBM29	A	8*	petite

Continued on next page

Table A.1 – continued from previous page

Date of isolation	CBM Line	Box	copper12 growth (# of 8)	YPG growth
30 Jan 2011	CBM30	B	5	✓
30 Jan 2011	CBM31	B	-	petite
31 Jan 2011	CBM32	B	-	petite
31 Jan 2011	CBM33	B	8	✓
31 Jan 2011	CBM34	B	8	✓
31 Jan 2011	CBM35	B	0	✓
31 Jan 2011	CBM36	B	8	✓
31 Jan 2011	CBM37	B	8	✓
31 Jan 2011	CBM38	C	-	petite
31 Jan 2011	CBM39	C	0	✓
2 Feb 2011	CBM40	A	-	petite
2 Feb 2011	CBM41	A	0	✓
2 Feb 2011	CBM42	A	-	petite
2 Feb 2011	CBM43	A	-	petite
2 Feb 2011	CBM44	B	8	✓
2 Feb 2011	CBM45	B	8	✓
3 Feb 2011	CBM46	A	4	✓
3 Feb 2011	CBM47	A	7	✓
3 Feb 2011	CBM48	A	-	petite
3 Feb 2011	CBM49	B	8	✓
3 Feb 2011	CBM50	B	0	✓
3 Feb 2011	CBM51	B	5	✓
3 Feb 2011	CBM52	B	0	✓
3 Feb 2011	CBM53	B	6	✓
3 Feb 2011	CBM54	B	3	✓
3 Feb 2011	CBM55	B	4	✓
3 Feb 2011	CBM56	C	no growth on YPD	-

* These lines were included in our study because colonies were not noticeably petite on YPD. Whole-genome sequencing indicated very little depth of coverage for mitochondrial genes for the copper resistant colonies analysed. These were subsequently shown to be incapable of growth on YPG plates (respiration deficient).

† Because colonies on YPD plates were not noticeably petite, this line was assayed for copper tolerance. As none of the 8 colonies grew, this line was dropped.

‡ These colonies were small on YPD but later shown to be capable of respiration (growth on YPG). Whole-genome sequencing was then conducted to determine the genetic basis of copper resistance.

Table A.2: T-test results comparing maximum growth rates of the CBM lines to growth of BY4741 in YPD + ferric citrate. Maximum growth rates were highly correlated among iron concentrations (10mM vs 40mM: $r = 0.89$, $t = 11.00$, $df = 31$, $p < 0.001$; 60mM vs 40mM: $r = 0.56$, $t = 3.72$, $df = 31$, $p = 0.0008$), so only 40mM results are presented in the text. Growth was assayed by automated OD readings over a 24 hour period in the Bioscreen C.

	10 mM ferric citrate			**40 mM ferric citrate**			**60 mM ferric citrate**		
Line	t	df	*p*-value	t	df	*p*-value	t	df	*p*-value
CBM1	-0.14	6.09	0.89	-0.34	7.80	0.74	-0.83	6.35	0.44
CBM2	-1.23	4.64	0.28	-5.47	6.78	0.001	1.8	4.83	0.13
CBM3	-0.04	5.75	0.97	-7.01	7.81	0.0001	2.81	6.6	0.03
CBM4	1.03	6.97	0.34	-2.16	5.43	0.078	0.32	6.23	0.76
CBM5	-1.86	4.69	0.13	-0.03	6.16	0.98	0.89	11.04	0.39
CBM6	-4.09	6.01	0.01	-8.48	5.86	0.0002	-2.59	9.58	0.03
CBM7	-10.42	5.64	0	-12.66	5.49	< 0.0001	-0.67	6.12	0.53
CBM11	-0.58	5.26	0.59	-9.81	4.99	0.0002	3.27	5.79	0.02
CBM13	-0.19	7.99	0.86	-0.22	5.80	0.83	1.89	4.67	0.12
CBM14	2.51	7.27	0.04	-0.0001	5.23	1.00	1.42	4.69	0.22
CBM16	-2.76	7.88	0.03	-27.41	7.61	< 0.0001	-2.58	8.05	0.03
CBM17	-8.88	4.52	0	-16.51	7.24	< 0.0001	-4.86	18.19	0
CBM18	-1.41	4.85	0.22	-3.42	6.07	0.01	1.31	4.5	0.25
CBM20	-6.3	4.42	0	-33.61	4.64	< 0.0001	-2.1	6.95	0.07
CBM21	-9.09	4.96	0	-11.87	9.00	< 0.0001	0.29	6.68	0.78
CBM22	-2.85	7.91	0.02	-11.10	5.11	0.0001	-3.74	8.64	0
CBM24	1.32	4.79	0.25	-1.17	5.32	0.29	0.34	6.71	0.75
CBM25	0.5	7.28	0.63	3.67	5.19	0.01	1.24	25	0.23
CBM26	-15.22	7.59	0	-36.20	5.11	< 0.0001	-3.01	12.37	0.01
CBM29	-4.92	7.69	0	-28.54	7.87	< 0.0001	-1.91	18.92	0.07
CBM30	-0.41	5.85	0.7	-6.37	7.97	0.0002	4.36	6.03	0
CBM33	-1.28	7.96	0.24	-1.49	4.83	0.20	1.59	6.38	0.16
CBM34	-3.07	7.92	0.02	-2.14	5.72	0.08	1.45	6.84	0.19
CBM36	-1.42	6.01	0.21	0.26	5.31	0.81	0.01	5.78	1
CBM37	0.06	5.45	0.96	-0.38	5.22	0.72	-1.37	8.43	0.21
CBM45	-0.26	5.69	0.8	-0.13	5.10	0.90	1.31	6.04	0.24
CBM46	-0.3	7.67	0.77	-0.90	4.69	0.41	0.89	6.23	0.41
CBM47	0.05	5.99	0.96	-0.82	4.95	0.45	2.45	13.5	0.03
CBM49	-11.04	4.48	0	-7.83	4.72	0.0007	-0.38	5.16	0.72
CBM51	0.44	7.68	0.67	-0.21	5.74	0.84	0.52	10.86	0.61
CBM53	-0.01	7.97	0.99	-0.93	5.36	0.39	0.44	5.54	0.68
CBM54	0.14	7.6	0.89	1.18	7.28	0.27	1.09	6.95	0.31
CBM55	1.38	4.5	0.23	1.54	7.84	0.16	2.12	7.33	0.07

Table A.3: Predicted transcription factor binding site gains and losses from intergenic mutations. The nearest ORFs upstream (5' on the Watson strand) or downstream (3') are given, as well as the distances to the start sites (brackets) and the orientation of the binding site (S: binding site precedes start of ORF; E: binding site after end of ORF). In bold are binding sites within 500 bp of the start of a gene on the coding strand. Neighbouring repeat elements, multi-copy tRNAs, or dubious ORFs were ignored.

Line	Position	Upstream ORF	Downstream ORF	Mutation	TF lost	TF gained
CBM1	XVI.420661	**YPL071C [143] (S)**	***MUK1* [287] (S)**	A>T	n/a	n/a
CBM3	VII.150650	***COX4* [479] (S)**	*TPN1* [2126] (E)	G>T	*FKH2*, *FKH1*, *HCM1*	*ORC2*, *SFP1*, *SPT15*
CBM5	XIV.284255	YNL190W [1860] (E)	***SRP1* [5] (S)**	T>G	*FKH2*, *HCM1*, *SUM1* *NHP6A*, *NHP6B*, *ORC2* *PHO2*, *SMP1*, *SPT15*, *YAP1*	n/a
CMB5,13,21	X.654261	**YJR124C [23] (S)**	*ENT3* [1702] (E)	T>C	*SUM1*, *ORC2*, *STB3*	n/a
CBM7	III.306327	YCR102C [860] (S)	*PAU3* [1474] (S)	G>T	n/a	*GAT1*, *GLN3*
CBM7	IX.370383	***PAN1* [475] (S)**	**YIR007W [321] (S)**	C>G	n/a	*GAT1*, *GLN3*, *GZF3*, *ECM23*, *SRD1*
CBM11	XII.605283	***CDC42* [496] (S)**	*BNA5* [1836] (E)	1D indel	n/a	*ORC2*, *SFP1*, *YGR067C*
CBM24	IV.805517	*SEC7* [3295] (S)	*HSP42* [1104] (S)	G>A	*STP4*	*UME6*
CBM24	IV.805485	*SEC7* [3263] (S)	*HSP42* [1136] (S)	A>G	*LYS14*, *YKL222C*, *YRR1*	*AR080*, *CEP3*, *PUT3*, *RDS2*, *TBS1*
CBM29	XV.566240	***ADE2* [49] (S)**	*AFI1* [3318] (E)	G>C	*RAP1*	n/a
CBM29	VII.1376	(telomere)	*COS12* [1414] (S)	A>C	*NHP6B*, *NHP6A*, *ORC2* *PHO2*, *SPT15*, *YOX1*	n/a
CBM34	XI.364516	*PTM1* [1894] (S)	*SNR69* [260] (S)	complex 1I indel	n/a	*GAT1*, *GLN3*, *GZF3*, *NHP6B*, *PHO2*, *SFP1*
CBM49	V.438349	YER134C [546] (S)	*GDI1* [1267] (S)	G>C	*SKN7*	n/a
CBM49	XIII.420239	***PDS5* [210] (S)**	*VPS20* [1910] (E)	A>C	*DOT6*	*ERT1*
CBM51	IV.310552	***RPP1A* [430] (S)**	*THI3* [1919] (E)	A>G	*FKH2*, *HCM1*, *NHP6A*, *NHP6B*, *PHO2*, *SPT15*	*ORC2*, *SUM1*

Table A.4: T-test results comparing maximum growth rate in copper8 between CBM lines and BY4741. Growth was assayed by automated OD readings over a 24 hour period in the Bioscreen C.

	Maximum growth rate		
Line	t	df	*p*-value
CBM1	13.65	25.16	< 0.0001
CBM2	11.41	24.19	< 0.0001
CBM3	10.60	25.60	< 0.0001
CBM4	9.32	25.73	< 0.0001
CBM5	14.47	24.05	< 0.0001
CBM6	13.25	24.39	< 0.0001
CBM7	8.80	24.11	< 0.0001
CBM11	12.78	24.79	< 0.0001
CBM13	14.36	24.56	< 0.0001
CBM14	11.09	25.54	< 0.0001
CBM16	11.44	24.71	< 0.0001
CBM17	12.09	24.68	< 0.0001
CBM18	15.29	25.43	< 0.0001
CBM20	11.56	24.36	< 0.0001
CBM21	8.12	24.19	< 0.0001
CBM22	12.21	25.66	< 0.0001
CBM24	0.43	5.49	0.68
CBM25	9.93	25.85	< 0.001
CBM26	3.81	27.14	0.0007
CBM28	6.73	25.66	< 0.0001
CBM29	11.02	24.65	< 0.0001
CBM30	10.72	25.89	< 0.0001
CBM33	14.71	24.24	< 0.0001
CBM34	11.10	26.13	< 0.0001
CBM36	12.53	21.12	< 0.0001
CBM37	14.89	27.70	< 0.0001
CBM44	14.02	26.02	< 0.0001
CBM45	13.05	25.99	< 0.0001
CBM46	13.78	26.47	< 0.0001
CBM47	14.37	27.20	< 0.0001
CBM49	7.88	26.01	< 0.0001
CBM51	7.32	25.05	< 0.0001
CBM53	10.24	25.84	< 0.0001
CBM54	4.54	24.60	0.0001
CBM55	11.30	26.39	< 0.0001

Table A.5: T-test results comparing maximum growth rate in YPD between CBM lines and BY4741. Growth was assayed by automated OD readings over a 24 hour period in the Bioscreen C.

	Maximum growth rate		
Line	t	df	*p*-value
CBM1	-2.62	27.98	0.014
CBM2	-1.50	6.89	0.18
CBM3	-4.72	26.93	<0.0001
CBM4	0.90	5.96	0.40
CBM5	-1.44	18.46	0.17
CBM6	-5.86	26.06	<0.0001
CBM7	-2.20	5.95	0.07
CBM11	-3.86	14.88	0.002
CBM13	-2.31	27.00	0.028
CBM14	0.0004	7.30	1.00
CBM16	-6.69	25.79	<0.0001
CBM17	-8.63	26.77	<0.0001
CBM18	-0.66	7.18	0.53
CBM20	-0.48	5.34	0.65
CBM21	-0.22	5.02	0.84
CBM22	-8.88	26.52	<0.0001
CBM24	1.86	8.33	0.10
CBM25	1.69	6.10	0.14
CBM26	-19.28	26.76	<0.0001
CBM28	-7.63	24.87	<0.0001
CBM29	0.24	5.65	0.82
CBM30	-1.80	6.04	0.12
CBM33	-1.63	9.12	0.14
CBM34	-3.89	26.23	0.0006
CBM36	-5.19	24.35	<0.0001
CBM37	-1.54	6.94	0.17
CBM44	-1.50	9.98	0.17
CBM45	-4.23	26.70	0.0002
CBM46	-4.47	25.03	0.0001
CBM47	-1.51	9.90	0.16
CBM49	-9.68	25.53	<0.0001
CBM51	0.20	6.21	0.85
CBM53	0.30	9.01	0.77
CBM54	0.46	6.10	0.66
CBM55	1.96	5.78	0.10

Table A.6: T-test results comparing maximum growth rate in YPG between CBM lines and BY4741. Growth was assayed by automated OD readings over a 24 hour period in the Bioscreen C.

	Maximum growth rate		
Line	t	df	*p*-value
CBM1	0.25	13.04	0.81
CBM2	-5.33	16.19	0.0001
CBM3	-6.68	18.30	< 0.0001
CBM4	1.01	17.52	0.33
CBM5	-0.30	14.22	0.77
CBM6	-56.27	17.09	< 0.0001
CBM7	-5.78	16.03	< 0.0001
CBM11	-8.83	15.96	< 0.0001
CBM13	-1.03	11.74	0.32
CBM14	0.84	13.59	0.42
CBM16	-59.60	18.40	< 0.0001
CBM17	-12.32	12.62	< 0.0001
CBM18	-0.41	21.13	0.69
CBM20	-69.59	24.25	< 0.0001
CBM21	-21.28	13.39	< 0.0001
CBM22	-31.41	20.28	< 0.0001
CBM24	-1.58	23.14	0.13
CBM25	1.94	19.23	0.07
CBM26	-21.35	20.68	< 0.0001
CBM28	-39.78	36.21	< 0.0001
CBM29	-91.71	45.36	< 0.0001
CBM30	-5.56	21.67	< 0.0001
CBM33	-1.64	26.01	0.11
CBM34	-0.75	15.98	0.46
CBM36	-0.20	13.43	0.85
CBM37	-1.33	14.82	0.20
CBM44	-1.71	15.21	0.11
CBM45	-2.14	12.96	0.05
CBM46	-1.04	16.19	0.31
CBM47	-0.35	13.78	0.73
CBM49	-7.57	13.07	< 0.0001
CBM51	0.27	14.87	0.79
CBM53	0.18	13.44	0.86
CBM54	-0.61	13.69	0.55
CBM55	-0.77	18.51	0.45

Table A.7: Summary of linear model analyses of the maximum growth rate of tetrads, assayed in YPD within the Bioscreen C. For each line, we show the analysis of a full model accounting for the genes listed below. Thus, CBM2 shows evidence for an effect of an additional copy of chrII on growth in YPD. Significant *p*-values are in bold.

Line	Term	Estimate	t	df	*p*-value
CBM2	*CUP1*	-4.66E-08	-0.48	13	0.64
	+chrII	-7.37E-02	-3.39	13	**0.0048**
CBM14	*CUP1*	8.20E-08	1.60	21	0.12
	MAM3	-3.61E-02	-1.75	21	0.095
CBM25	*CUP1*	2.71E-08	0.62	12	0.54
	MLP1	-2.67E-02	-1.086	12	0.30
	ENA5	5.62E-03	0.26	12	0.80
CBM34	*CUP1*	-1.58E-08	-0.48	21	0.63
	VTC4	2.17E-02	1.29	21	0.21

Table A.8: Summary of linear model analyses of the maximum growth rate of tetrads assayed in copper9, after correcting for growth in YPD (maximum growth rate in copper9 minus maximum growth rate in YPD), both measured within the Bioscreen C. For each line, we show the analysis of a full model accounting for the genes listed below. Only CBM2 showed evidence for an effect of a mutation on growth in YPD (see Table A.7). Significant *p*-values are in bold.

Line	Term	Estimate	t	df	*p*-value
CBM2	*CUP1*	2.96E-07	3.18	13	**0.0072**
	chrII	1.10E-01	5.30	13	**0.00015**
CBM14	*CUP1*	4.40E-08	0.80	21	0.43
	MAM3	5.49E-02	2.48	21	**0.022**
CBM25	*CUP1*	8.50E-08	1.53	12	0.15
	MLP1	2.11E-02	0.67	12	0.51
	ENA5	-5.93E-03	-0.21	12	0.84
CBM34	*CUP1*	1.70E-07	4.12	21	**0.00049**
	VTC4	3.48E-02	1.65	21	0.11

Table A.9: Additional mutations identified in the small-colony forming CBM lines. *CUP1* coverage for each line is provided in the second column and does not account for additional copies via chrVIII aneuploidy.

CBM line	*CUP1* coverage	Genome Position (chr.bp)	Gene	Mutation (Watson strand)	Position (from 5' end)	Amino acid change	Exchangeability
CBM9	0.69	VII.481622	*PMA1*[a]	C>T	1045	Gly>Ser	0.304
CBM27	0.83	VII.482121	*PMA1*	3D indel (AAC/—)	544	Val> −	
		mito.83071	intergenic	A>G			
CBM28	1.25	IV.43829&IV.43830	intergenic	CA>AT			
		chrIII aneuploidy					
		chrV aneuploidy					
		chrVIII aneuploidy					

[a] As a sample from the population was sequenced, this mutation was not fixed but was called as a "heterozygote" (43.4% of reads).

A.3 Supporting Figures

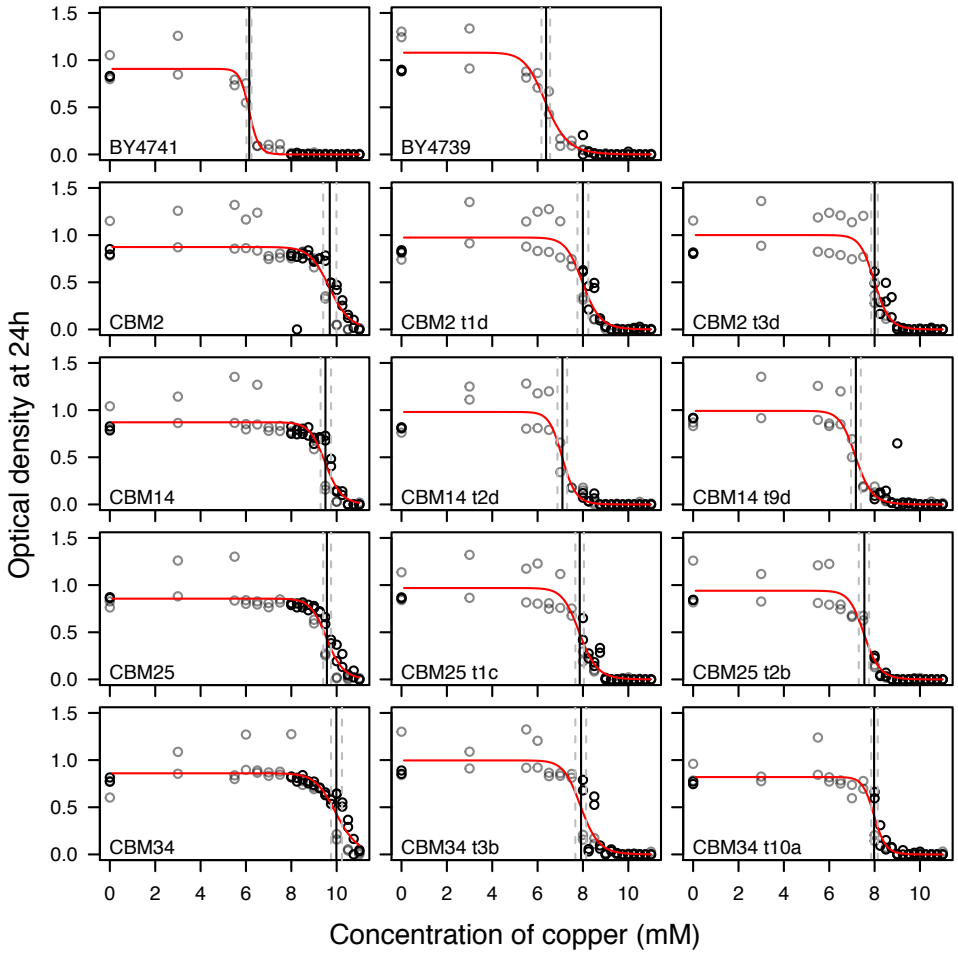

Figure A.1: Optical density after 24 hours of growth in the Bioscreen C for specific spores over a range of copper concentrations. Spores were chosen that had lower *CUP1* copy number and carried either an extra copy of chrII (CBM2 lines), a SNP in *MAM3* (CBM14 lines), a SNP in *MLP1* (CBM25 lines) or a SNP in *VTC4* (CBM34 lines). Grey and black circles represent data points collected on two separate days, with two replicates per day. Curves drawn in red are maximum likelihood fits using the methods described in GERSTEIN *et al.* (2012), with the estimated IC_{50} represented by a vertical black line and its corresponding 95% confidence interval shown by the grey dashed lines.

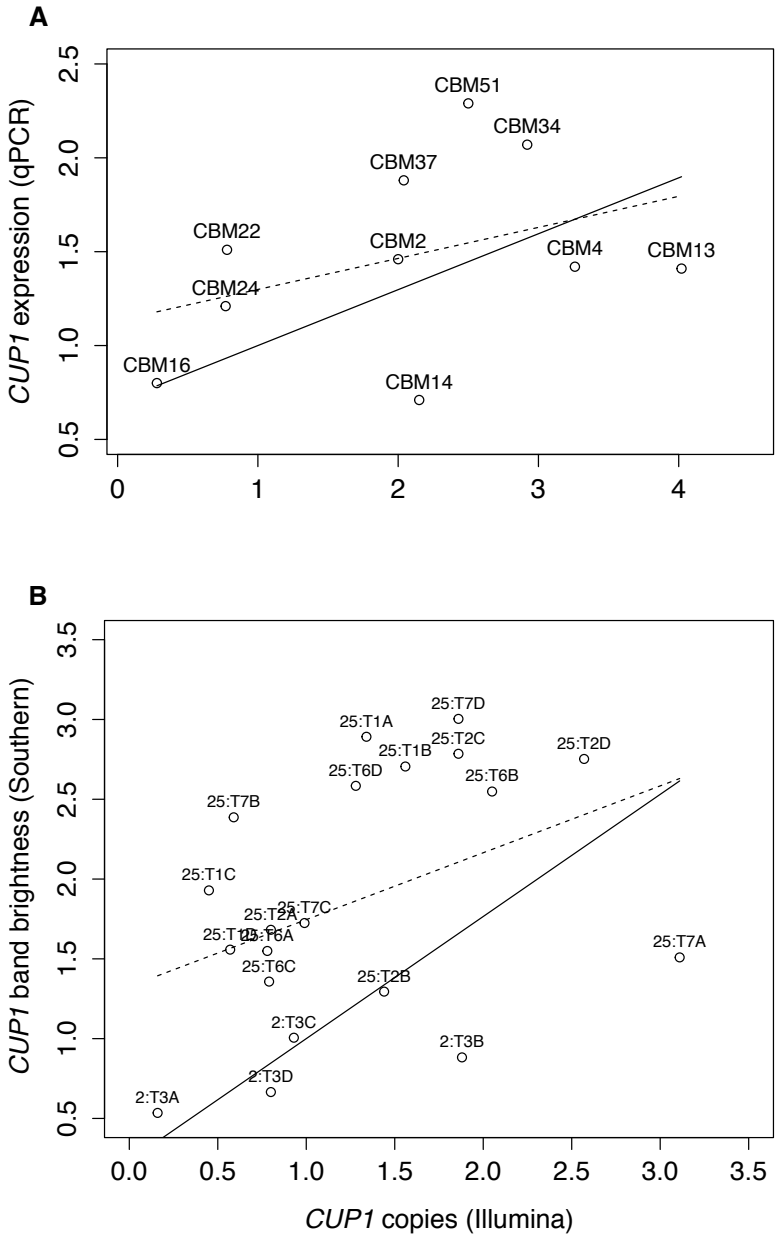

Figure A.2: Comparison of *CUP1* copy number assays. A. *CUP1* expression level was determined via qPCR and compared with the *in silico* qPCR estimates based on the fastq Illumina files for ten CBM lines. Expression levels for *CUP1* (normalized to *TAF10*) were obtained by qPCR, with the y-axis giving expression levels relative to the BY4741 ancestral line. The slope is significant when forced through the (1,1) point, which assumes that both axes are scaled to the ancestor (even though the derived BMN lines and not BY4741 were used as the control in the *in silico* qPCR assays; p = 0.02, solid), but the slope is not significant otherwise (p = 0.27, dashed). B. *CUP1* copy number was estimated by band brightness from Southern blots and compared with *in silico* qPCR estimates for the lines established from tetrads. Band brightness from the Southerns was normalized to the average of two BY4739 bands run on the same gels (recall that the CBM lines had been crossed to BY4739 to generate the tetrads; normalizing to the two BY4741 bands yielded similar results). Only those tetrads for which whole-genome sequencing was performed are included (e.g., "25:t1a" refers to "CBM25, tetrad 1, colony a"). The slope is significant when forced through the (1,1) point (p = 0.02; solid) and marginally significant otherwise (p = 0.08, dashed).

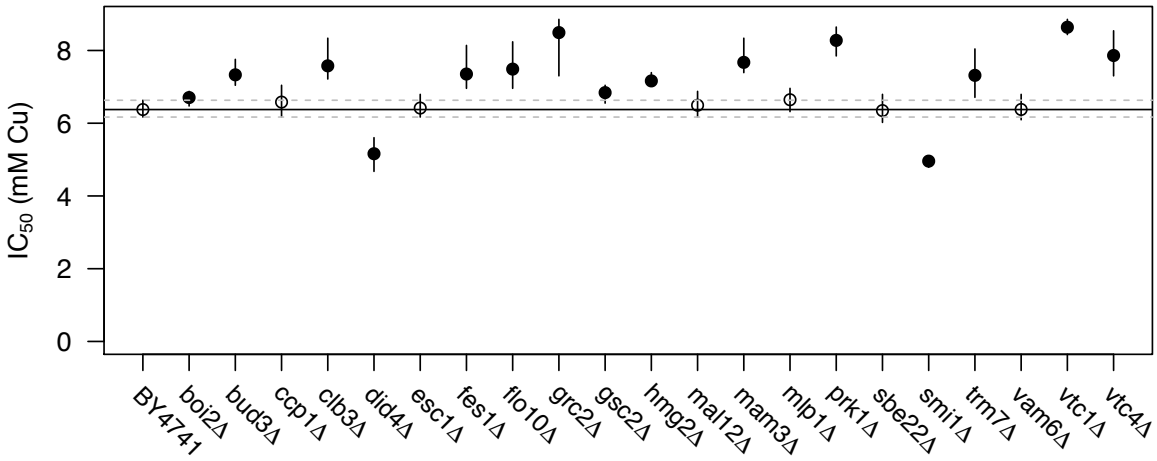

Figure A.3: Copper tolerance of *S. cerevisiae* knockout lines for genes identified in our experiment. Solid circles identify lines that have significantly different tolerance than BY4741, measured as IC_{50} (bars represent 95% confidence intervals). The horizontal lines are for illustrative purposes to indicate the mean (solid line) and confidence interval (dashed lines) for BY4741.

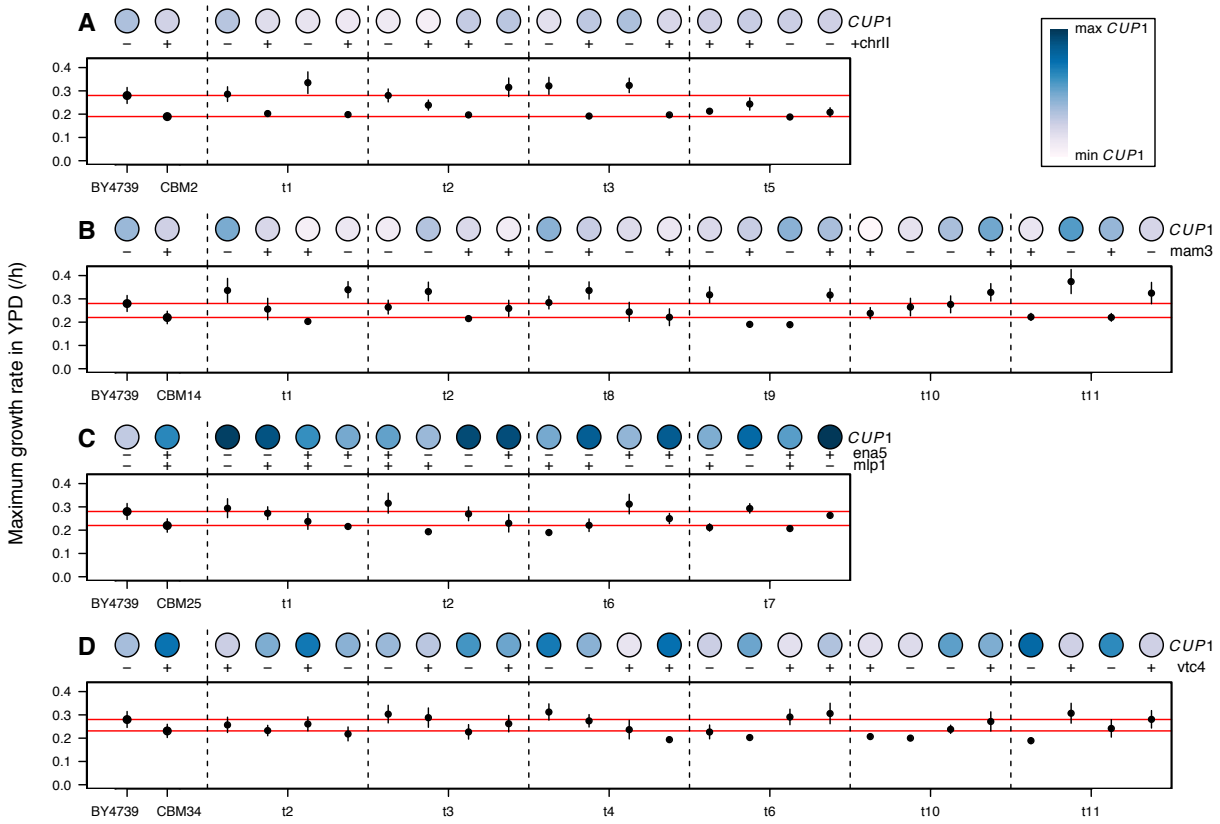

Figure A.4: Maximum growth rate of tetrads in YPD, as measured by the Bioscreen C. Tetrads were derived from four different CBM lines: A. CBM2, B. CBM14, C. CBM25, and D. CBM34. For each line, maximum growth rate was assayed within the Bioscreen C on a single day (± 1 SE across replicate wells). The darkness of the circle represents the relative number of copies of *CUP1*, as assayed from Southern blots. Presence ($+$) or absence ($-$) of a segregating mutation is also noted. All lines are compared to the growth rate of the two parents, BY4739, and the relevant CBM parent (red lines), except for the tetrads derived from CBM25 for which parental growth rate was not assayed (due to its initially being considered CBM22, see Materials and Methods).

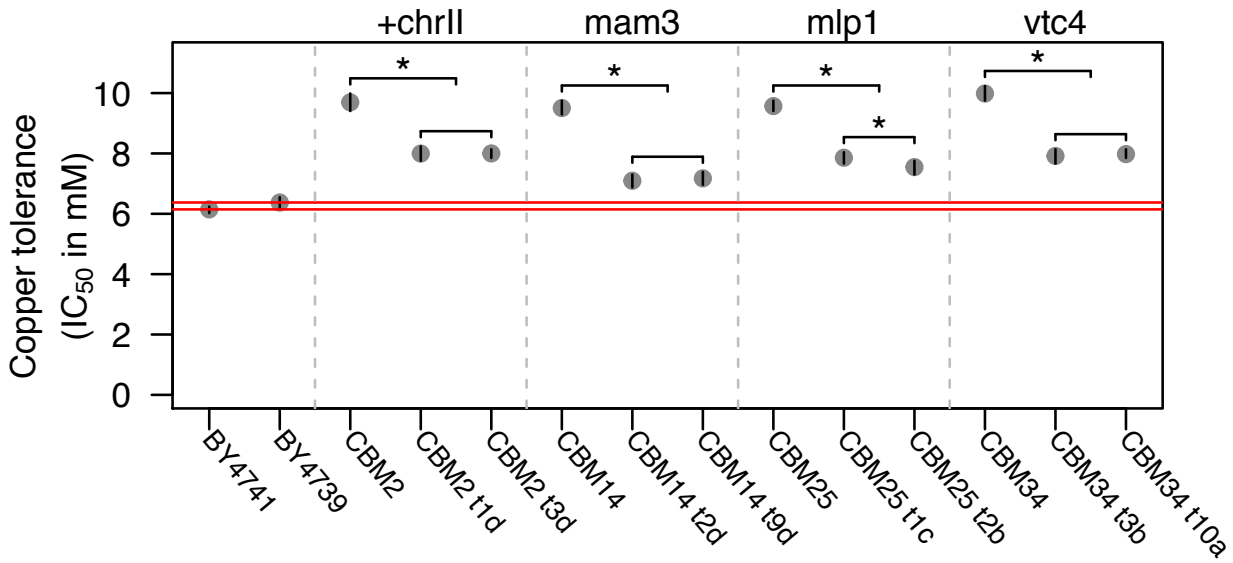

Figure A.5: Copper tolerance, as measured by IC_{50} after 24 hours of growth in the Bioscreen C for specific spores. Lines were chosen because they had low *CUP1* copy number and carried either an extra copy of chrII (CBM2 lines), a mutation in *MAM3* (CBM14 lines), a mutation in *MLP1* (CBM25 lines), or a mutation in *VTC4* (CBM34 lines). All mutant lines had a significantly higher IC_{50} than either of the BY controls, and all spores had a significantly lower IC_{50} than their CBM parent. Horizontal bars indicate statistical comparisons, where an asterisk (*) above a bar indicates statistical significance ($p < 0.05$). Among the spores carrying the same allele, only the CBM25 spores differed significantly from one another in IC_{50}, and only marginally so if corrected for multiple comparisons. Note that CBM25 t1c also carries the mutation in *ENA5*. Vertical bars represent 95% confidence intervals.

Appendix B

Appendix for Chapter 3: Widespread Genetic Incompatibilities Between First-Step Mutations During Parallel Adaptation of *Saccharomyces cerevisiae* to a Common Environment

B.1 Supplementary Methods

B.1.1 Strain construction details

All possible haploid and diploid genotypes were created for each pair of four beneficial mutations (one in each of *ERG3, ERG5, ERG6* and *ERG7*, Table 3.1). Each mutation was initially isolated in the BY4741 haploid background (*MATa his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0*) and given a Beneficial Mutation Nystatin (BMN) strain number (GERSTEIN *et al.* 2012). Each BMN strain was mated to BY4739 (*MAT*α *leu2*Δ*0 lys2*Δ*0 ura3*Δ*0*) (Open Biosystems) to create strains heterozygous for a single ERG mutation, and diploids were positively selected on plates lacking both histidine and lysine. Similarly, diploid non-mutant strains were created by mating BY4741 and BY4739. In each case, single colonies were then grown up on a second selection plate and frozen at -80°C in 15% glycerol.

MATα single mutant strains were isolated by sporulation of the heterozygous diploids. Diploid stock grown on a YPD plate was used to inoculate 10 mL of YPD and grown overnight on a rotor at 30°C. 200 μL of culture was then washed, spread on potassium acetate plates (1% KOAc, 2% agar) and sporulated at 25°C until a sufficient number of tetrads could be observed. The resulting tetrads were dissected by micromanipulation on YPD plates. The spores were allowed to germinate and grown at 30°C for three days before replica plating to test for auxotrophies, mating type, and nystatin growth ability. Auxotrophy was assessed on SC plates lacking the appropriate amino acid. Mating type was tested by replica plating tetrads onto plates containing a lawn of *MATa* or *MAT*α yeast carrying a histidine (*his1-123*) auxotrophy, allowing them to mate, and subsequently testing for mating on a plate lacking arginine, histidine, leucine, lysine, methionine, tryptophan, adenine and uracil (i.e., a plate on which no original haploid strain could grow). Nystatin growth was assessed on YPD + 8 μM nystatin because growth of ancestral strains was not noticeably inhibited on plates with a lower concentration of nystatin. YPD + 8 μM nystatin plates

were made by preparing YPD medium with agar as usual, subsequently adding the appropriate amount of 2.7 mM nystatin stock, and mixing by inversion immediately before pouring. All tetrads were verified for 2:2 segregation of auxotrophies and mating type. Once this was confirmed, the spores that showed growth on the nystatin plate and contained the desired *MAT*α *lys2*Δ*0* mutation were frozen at -80°C in 15% glycerol. Throughout strain construction, histidine and lysine auxotrophies were consistently kept with the same mating types so that all haploid strains were either *MATa his3*Δ*1* or *MAT*α *lys2*Δ*0*. The methionine auxotrophy (*met15*Δ*0*) did not show strong selection on plates lacking methionine and was not tracked.

The haploid *MAT*α strains were then mated to the original *MATa* strains to create strains that were either homozygous for one mutation or heterozygous for two mutations. Diploids were selected and frozen as described for the singly heterozygous strains.

The haploid double mutant strains were created through sporulation and dissection of the double heterozygous strains. Three strains (*erg3/ERG3 erg6/ERG6*, *erg3/ERG3 erg7/ERG7*, and *erg6/ERG6 erg7/ERG7*) were struck from frozen on YPD plates and grown at 30°C for 2-3 days. They were sporulated, dissected and checked as described above except that they were moved to 20°C after three days of sporulation. This protocol was repeated for *erg3/ERG3 erg5/ERG5*, except that it was kept at 20°C from the beginning. *erg5/ERG5 erg7/ERG7* would not sporulate under these conditions. To obtain the *MATa* double mutant strain it was sporulated in 10 mL 1% KOAc + amino acids liquid medium at 20°C. In order to obtain the *MAT*α double mutant strain, it was sporulated by streaking a patch of cells onto a GNA pre-sporulation plate (5% dextrose, 3% nutrient broth, 1% yeast extract, 2% agar) and growing at 30°C overnight, repeating the streaking and growth on another GNA pre-sporulation plate, and finally sporulating in 2 mL of supplemented sporulation medium (1% potassium acetate, 0.005% zinc sulphate, 2 mg/100 ml uracil, 10 mg/100 ml leucine) on a rotor at 25°C for five days, then moving to 30°C until a sufficient number of tetrads were found. The same procedure was applied to *erg5/ERG5 erg6/ERG6* to obtain the *MAT*α double mutant strain. Tetrads were chosen that showed 2:2 segregation of the nystatin resistance (assessed either on YPD + 8 μM nystatin plates, YPD + 10 μM nystatin plates or in a liquid assay), indicative of two double mutant spores and two wildtype spores. Double mutant strains were frozen at -80°C in 15% glycerol. All haploid double mutant strains were confirmed by Sanger sequencing.

We failed to obtain the *MATa erg5 erg6* double mutant strain through crossing and sporulation because the two genes are linked (they are 48 kb apart but flank the centromere of chr XIII). For this strain, transformations were performed using a protocol based on CREGG (2007). *MATa erg5* yeast were grown from a single colony in 10 mL YPD at 30°C. The next day, two new 10 mL YPD tubes were inoculated with 500 μL of yeast from the overnight culture and grown at 30°C until reaching an OD_{600} between 0.5 and 0.6. One tube was used for the transformation, and one was used as a negative control. Cells were collected by spinning the cultures down for 5 minutes at 4500 rpm and were washed twice with water using a spin of 10 minutes at 4500 rpm. The yeast were resuspended in 2 mL of cold 1 M sorbitol, spun at 5000 rpm long enough to pellet the cells, the supernatant was removed, and the yeast were resuspended in 1 mL of cold 1 M sorbitol. 80 μL of these cultures were then electroporated, along with either 8 μg of an oligonucleotide designed to contain the *ERG6* SNP of interest (sequence: TTCAAAGAGGCGATTTAGTTCTCGACGTTCGTTGTG-GTGTTGGGGGCCCAGCAAG) or an equal volume of water, using a BioRad Gene Pulser Xcell and the

parameters defined in CREGG (2007). Immediately after electroporation, 1 mL of YPD was added to the yeast and the cells were incubated for 1 hour at 30°C to recover. The cells were then plated on YPD + 10 μM nystatin plates and incubated at 30°C until colonies were visible. The insertion of the mutation in *erg6* was verified by Sanger sequencing.

Strains with one heterozygous and one homozygous locus as well as double homozygous strains were created by mating the corresponding single mutant strains or *MATa* double mutant strains to the *MAT*α double mutant strains, as described above.

B.1.2 Segregating mutation in *DSC2*

The original strain with a mutation in *ERG7* also carried a second mutation in the gene *DSC2* (Table 3.1). This mutation was not originally tracked when constructing the strains and it was later identified by Sanger sequencing in all haploid strains constructed from the original strain carrying a mutation in *ERG7* (Table B.1). Two combinations of strains that differed in their status at *DSC2* between the mating types (*erg7* and *erg5 erg7*) and were tested for differences in maximum growth. No significant difference was found for growth rate in nystatin whether we treat each replicate as independent or average data points collected on the same day (Welch two sample t-tests with replicates treated as independent, *erg7*: t = -0.38, df = 49.56, P = 0.71; *erg5 erg7*: t = -1.01, df = 37.66, P = 0.32). In YPD, a significant difference was found only for the gene combination *erg5 erg7* and only when all replicates were treated as independent (Welch two sample t-tests, *erg7*: t = 0.27, df = 50, P = 0.79; *erg5 erg7*: t = -2.32, df = 35.97, P = 0.026). The test was not significant when data points for each day were averaged (*erg5 erg7*: t = -1.90, df = 7.87, P = 0.094). Furthermore, the difference between mutant and wildtype *DSC2* growth rates was in each case minor and did not substantially alter the data points illustrated in main text Fig 3.3 or the conclusions drawn.

Table B.1: *DSC2* allele status in haploid strains constructed from the original strain carrying a mutation in *ERG7*.

Strain	Mating Type	Allele status at *DSC2*
erg7	a	mutant
erg7	alpha	wildtype
erg3 erg7	a	mutant
erg3 erg7	alpha	mutant
erg5 erg7	a	mutant
erg5 erg7	alpha	wildtype
erg6 erg7	a	wildtype
erg6 erg7	alpha	wildtype

B.1.3 Preparing stocks for growth rate assays

A total of seven growth rate assays were conducted for our analysis of epistasis. We had originally intended to perform three assays, but four more were performed to maintain the intended level of replication after

encountering problems with growth and strain construction. For an overview of which lines were included in which fitness assays, see S1 Table and for complete information about the growth assays, see files deposited at Dryad. Many of the lines involved in this study had poor growth even in a rich medium. Because of this, care was taken to standardize initial cell densities ("pre-assays") for use in subsequent growth rate assays. General methods will be explained first, with exceptions to these methods explained subsequently.

The pre-assay took place in 100-well honeycomb Bioscreen plates using a permissive medium of 148.5 μL of YPD + 0.5 μM nystatin (except for the first assay, which used only YPD). YPD + 0.5 μM nystatin was used to help prevent reversion of strains with severe growth defects in YPD while still permitting the growth of all strains. The wells were inoculated with 1.5 μL of frozen culture. Replicates were randomized within plates, always including all lines on the same plate for a given pair of mutations. The plates were incubated in the Bioscreen C Microbiological Workstation at 30°C with maximum continuous shaking, measuring the optical density (OD) of the cultures every 30 minutes using the wideband filter.The cultures were incubated in this way for 72 hours, which was enough time for most strains to obtain clear growth (defined as a maximum OD of about two times the initial OD); anything below this threshold was excluded from analysis unless otherwise noted. Maximum OD was used to determine the volume to transfer for the growth rate assays. If it was above 1, we transferred 1.5 μL into one plate containing 148.5 μL of YPD and one plate containing 148.5 μL of nystatin2 (using the same randomized well map). If the maximum OD was below 1, it was rounded to the nearest 0.05, and the transferred volume was scaled accordingly (giving final volumes ranging between 150 μL and 156 μL).

To investigate whether the pre-growth medium influenced growth rate, we ran a sign test comparing the mean maximum growth rates in the nystatin2 assay between Assay 1 (in which all strains were pre-grown in YPD) and Assay 2 (in which all strains were pre-grown in 0.5 μM nystatin). All 47 strains that were included in both assays (and not omitted due to growth problems) were included in the sign test, which was run using the function *binom.test* in the package *stats* by counting the number of strains for which maximum growth rate was higher in Assay 2 and comparing that to what is expected by chance (p = 0.5). No significant difference was found ($P = 0.56$; similar results were obtained with a paired t-test: $P = 0.24$).

The *erg6/erg6 erg7/erg7* diploid strain showed consistently poor growth, and all of the data for this strain comes from the fourth and sixth assays where the pre-assay was conducted over a longer period of time in a larger volume of liquid in an attempt to initiate the assays with the same number of cells. Briefly, 10 mL of 0.5 μM nystatin in a test tube was inoculated with 15 μL of *erg6/erg6 erg7/erg7* from frozen two days before all other lines were inoculated from frozen. The tube was incubated at 30°C on a rotor for this time. On the day when all other strains were being inoculated from frozen, the 10 mL tube of *erg6/erg6 erg7/erg7* was spun down in multiple 1.5 mL tubes and concentrated into 500 μL in one tube. 150 μL of this concentrated culture was used to fill the appropriate wells of the pre-assay plate. Despite this extra growth time and concentrating of cells, *erg6/erg6 erg7/erg7* still did not grow to an OD above the threshold at the end of the pre-assay in one of the two cases where growth rate was measured for this line and only barely did so in the other. Yeast was added to the assay plates from these wells according to their measured OD after the growth phase even though the OD was below the threshold (up to 7.5 μL was transferred).

We also modified growth conditions for three other strains that showed poor growth in early pre-assays

(*MATa erg6 erg7*, *MATα erg6 erg7*, and *erg6/erg6*). Once low growth from frozen was established, 2 μL (rather than 1.5 μL) of frozen stock was used to inoculate the wells in the pre-assay plates. Backup tubes were also grown for these strains and described when used. In all cases, backup tubes that contained YPD + 0.5 μM nystatin as the growth medium were inoculated from frozen at the same time as the pre-assay plates and were incubated at 30 °C, shaking at 200 rpm.

In the pre-assay for the second growth rate assay, two out of four replicates of *MATα erg6 erg7* had still not grown to an OD above the threshold by 72 hours. One well was omitted. For the other well, 1 mL from a 10 mL backup tube (originally inoculated with 10 μL of frozen culture) was spun down at 3000 rpm for 3 minutes, and this concentrated culture was used to replace the 150 μL on the growth plate. New OD readings were taken, and the new OD was within the range measured for the other strains.

In the third pre-assay, four out of four *MATα erg6 erg7* wells were below the threshold for detecting growth after 72 hours. The liquid from the wells was replaced with culture from four 10 mL backup tubes (originally inoculated with 10 μL of frozen culture). After measuring the OD of these wells, one well was still not above the threshold; to ensure that enough cells were transferred for that one line, we concentrated the cells found in 1 mL of the culture from the corresponding tube by spinning them down using a tabletop centrifuge and removing most of the supernatant, leaving ~200 μL of concentrated culture. 1.5 μL of this culture was transferred directly to the honeycomb plate for the growth assay.

In the sixth pre-assay, one replicate of *MATα erg6 erg7* remained below the threshold for growth after 72 hours. The liquid from the well was replaced with culture from a 1.5 mL backup tube containing 500 μL of culture (originally inoculated with 5 μL of frozen culture). A new OD reading was taken of that well and was within the range measured for the other strains.

Following each pre-assay, growth rate assays were conducted in both YPD and YPD + 2 μM nystatin ('nystatin2'), as described in the main text.

B.1.4 Analysis including outliers

All qualitative relationships between strains and the main conclusions were insensitive to the exclusion or inclusion of the identified outliers, with two exceptions for the haploids in nystatin2 (see Fig B.6 and Fig B.7 for versions of Fig 3.3 and Fig 3.4 that include all outliers). One exception is that the *erg3 erg5* strain no longer had a significantly lower maximum growth rate than the *erg3* strain in nystatin2. This was due to one large outlier in the *MATa erg3* data, which exhibited almost no growth (maximum growth rate of 0.038), while all remaining wells (including both mating types) showed substantial growth (maximal growth rate ranged from 0.16 to 0.25 across 35 wells). The exclusion of this single outlier leads to the observation of a significant difference between the aforementioned strains.

A similar failure of one well to show substantial growth was observed in *erg6/erg6* and *erg3/erg3 erg6/ERG6*. In addition, two wells of *erg6/ERG6* showed substantially higher growth (0.17 and 0.19), compared to all remaining wells (0.0022 to 0.072 across 22 wells), although our outlier exclusion algorithm only allowed one point to be excluded per strain. These other examples did not affect the statistical results but suggest either occasional contamination or mutation.

The other statistical difference is that the *erg3 erg6* strain no longer had a significantly lower maximum

growth rate than the *erg6* strain in nystatin2. The difference between these strains is only slightly significant in the model excluding all outliers ($P = 0.047$) and becomes marginal when either including all outliers ($P = 0.083$) or excluding only the one outlier replicate of *erg3* ($P = 0.058$). We believe that this represents a lack of power to detect a true, small difference in the haploids as this relationship is supported in the homozygous diploids (excluding outliers: $P = 0.0069$; including outliers: $P = 0.041$). For the full alternative analysis without outlier removal, see ONO *et al.* (2016).

B.2 Supporting Table

Table B.2: Experimental design of the growth rate assays. In each epistasis assay, growth rate was measured in a Bioscreen C over a 24 hour period for a pair of ergosterol mutations (first column) using two replicate wells for each genotype (ancestral, single mutant, double mutant for haploids, including those heterozygous or homozygous for diploids), with the exception of double mutant haploids, which were measured in four replicate wells. Checkmarks indicate that all data from this assay was used while bullets indicate that some strains were omitted (see footnotes).

Mutation Pair	Assay 1	Assay 2[a]	Assay 3	Assay 4	Assay 5	Assay 6	Assay 7
erg3 & erg5	✓	✓	✓				
erg3 & erg6	✓	✓	✓				
erg3 & erg7	✓	✓	✓				
erg5 & erg6	*MAT*α only[b]	*MAT*α only[b]	*MAT*α only[b]	•[c]	•[c]	•[c]	✓[d]
erg5 & erg7	•[e]	•[e]	•[e]	✓	[f]	✓	
erg6 & erg7	•[g]	•[g]	•[g]	✓	[h]	✓	

[a] The Bioscreen bulb burned out during the pre-assay; OD readings taken after replacing the bulb were used to estimate the final cell densities for inoculation of the growth rate assays.

[b] Others not assayed because *MATa erg5 erg6* was unavailable.

[c] *erg5/erg5 erg6/erg6* data from this assay were omitted from final analyses (see [d]).

[d] Because the double homozygous mutant displayed high levels of growth in both YPD and nystatin, we were concerned that the stock might contain a mixture of resistant and non-resistant cells. We thus struck this stock down to colonies on a YPD plate, picked five colonies, and used the resulting five stocks to assay growth. Each stock was confirmed to be *erg5/erg5 erg6/erg6* by Sanger sequencing. For each of these stocks, two replicate wells (1 stock) or three wells (4 stocks) were used to measure growth. All mutant and non-mutant combinations of *erg5* and *erg6* (haploid and diploid) were also regrown in two replicate wells in this assay. The growth rates of the diploid double mutants did not differ qualitatively from previous assays in nystatin but were substantially lower in YPD, consistent with the population analyzed in previous assays having been polymorphic (allowing non-resistant cells to proliferate). Thus, only data from Assay 7 was used for the growth rate of *erg5/erg5 erg6/erg6*, although the qualitative results in nystatin are unaffected if all data were used. Data from each well were treated independently in the analysis, given that each was grown separately from frozen.

[e] *erg5/erg5 erg7/ERG7* data from this assay were omitted from final analyses due to an error in the creation of the original line (identified by Sanger sequencing).

[f] Data not collected due to Bioscreen machine error resulting in lower replication for the line *erg5/erg5 erg7/ERG7*.

[g] *erg6/erg6 erg7/erg7* and some *erg6 erg7* haploid data could not be used due to insufficient starting cell densities of the double mutant lines from the pre-assays so that appreciable growth was never observed.

[h] Data not collected due to machine error. Epistasis was large and easy to detect for this gene pair, despite the lower replication.

B.3 Supporting Figures

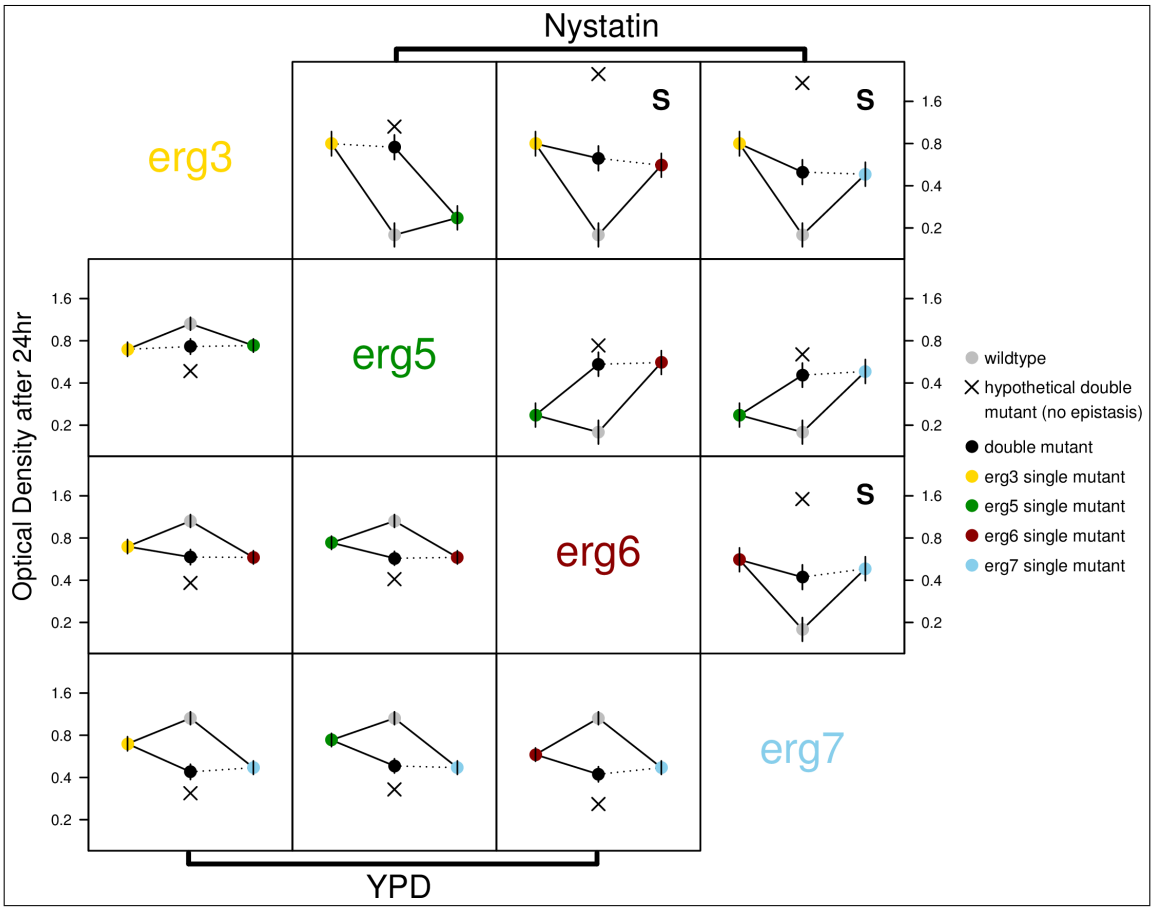

Figure B.1: Optical density after 24 hours of growth for haploid strains in nystatin2 (above diagonal) and YPD (below diagonal), plotted on a log scale. Points are the fitted least-squares means of the ODs, determined in the mixed-effects model run using log(OD). ×'s denote the additive fitness null expectation for the double mutant, i.e., with no epistasis. Each single mutant is coloured differently, the double mutant is shown in black, and the ancestor is grey. Vertical bars represent 95% confidence intervals of the fitted least-squares mean. Solid lines indicate significant comparisons, while dotted lines are non-significant comparisons. Combinations showing significant sign (S) and reciprocal sign (RS) epistasis are indicated by the presence of the abbreviation at the top of the panel. The same outliers were removed as in the analysis of maximum growth rate because their growth rates indicate a potential problem with the replicate. Sign epistasis is less often detected in this analysis of log(OD) in nystatin, likely because even slower growing strains are given time to catch up in cell density over 24 hours. All underlying raw data and analyses can be found in Ono *et al.* (2016).

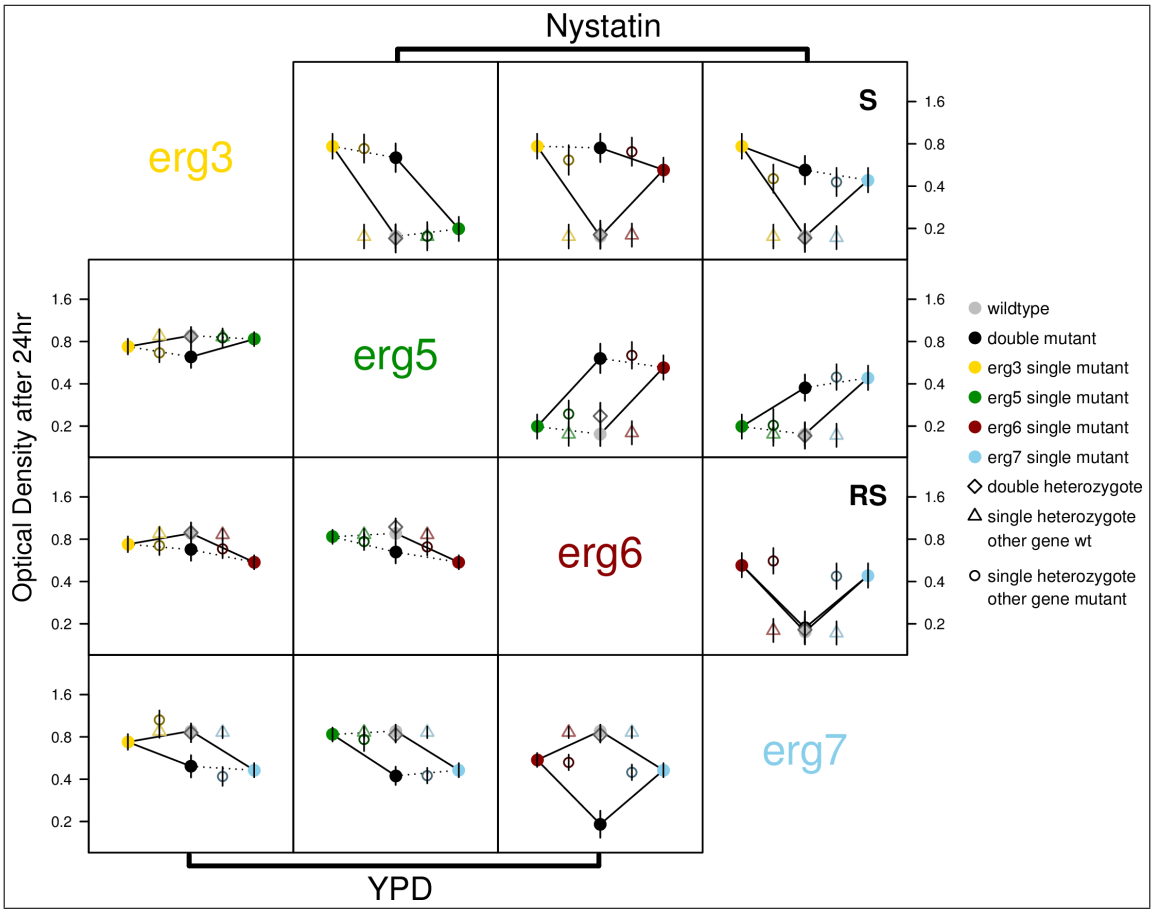

Figure B.2: Optical density after 24 hours of growth for diploid strains in nystatin2 (above diagonal) and YPD (below diagonal), plotted on a log scale. Points are the fitted least-squares means of the ODs, with closed circles determined in the mixed-effects model run using log(OD) including only homozygous strains and open symbols from the model that includes heterozygous strains (open diamonds: double heterozygotes; open triangles: single heterozygotes that are wildtype at the other gene; open circles: single heterozygotes that are homozygous mutants at the other gene). Points and bars are otherwise as in Fig B.1. All symbols are coloured intermediately according to genotype and arrayed along the x-axis so as to lie between the two strains that are genotypically most similar to it. Solid lines indicate significant comparisons in tests run including only homozygous strains while dotted lines are non-significant comparisons. See Fig B.1 for further graphical details. The same outliers were removed as in the analysis of maximum growth rate because their growth rates indicate a potential problem with the replicate. Sign epistasis is less often detected in this analysis of log(OD) in nystatin, likely because even slower growing strains are given time to catch up in cell density over 24 hours. Note that the strain *erg5/ERG5 erg6/erg6* was later found to be homozygous for the mutation in *ERG5*, likely due to a loss of heterozygosity event. All underlying raw data and analyses can be found in Ono *et al.* (2016).

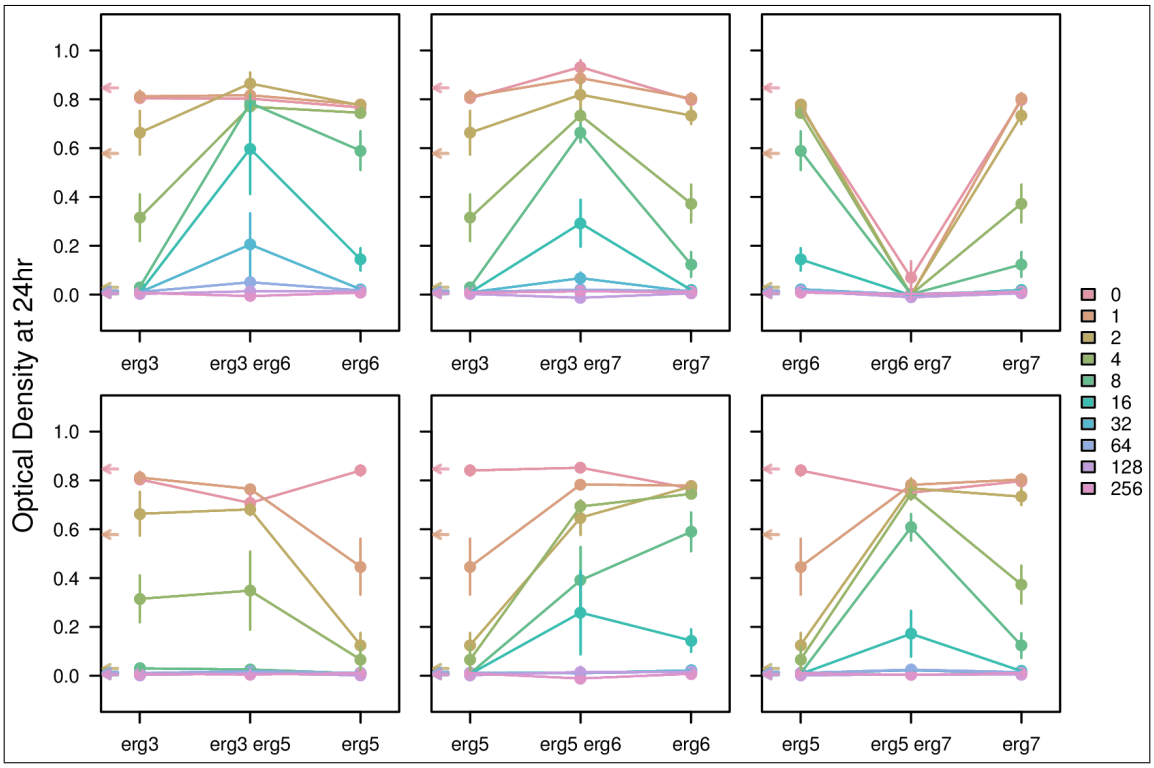

Figure B.3: Optical density after 24 hours of growth for homozygous diploids in a range of concentrations of nystatin. These results are qualitatively similar to the haploid strains with the exception of the *erg6/erg6 erg7/erg7* double mutant, which has very low growth in all concentrations of nystatin. Colours go from red to purple, through blues, from lowest to highest concentrations of nystatin. Lines connect different mutants in the same concentration of nystatin. Differences in OD between mutants were not tested statistically and are all represented by solid lines (in contrast to Fig 3.5). Arrows on the y-axes indicate the OD of the ancestral strain. All replicates were averaged, and error bars denote the standard error. Note that tolerance was assayed in the *erg5/erg5 erg6/erg6* homozygous double mutant before we determined that it was likely polymorphic; these points may thus be underestimates (see Table B.2 for details). All underlying raw data and analyses can be found in Ono *et al.* (2016).

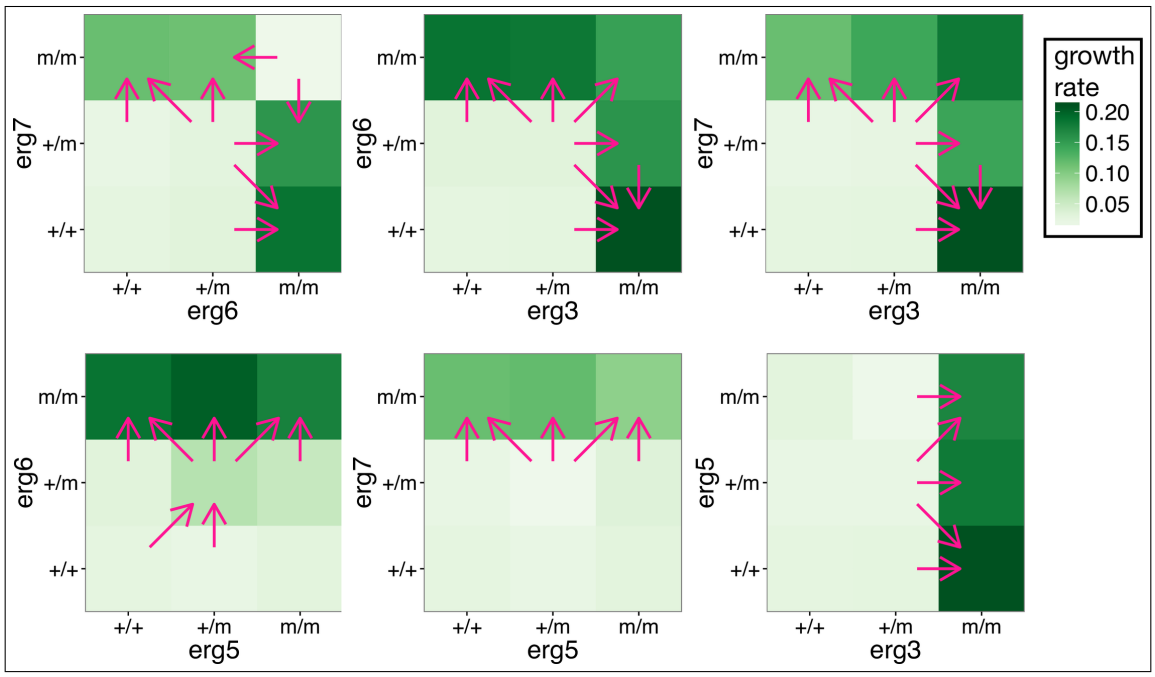

Figure B.4: Maximum growth rate of diploid strains for each gene combination in nystatin2. Genotype at each of the two genes combined is represented along the x- and y-axes, with the ancestral genotype in the lower left corner and the homozygous double mutant genotype in the upper right corner. Least-squares means of maximum growth rates, as determined from a model including all possible diploid genotypes, are represented by the darkness of the boxes. Arrows indicate significant differences between genotypes, with arrowheads pointing to the significantly higher growth rate as determined by pairwise comparisons corrected for multiple comparisons using the multivariate *t* distribution in *lsmeans*, as was done for the haploids and homozygous diploids. Only adjacent genotypes on the grid (horizontal and vertical) were compared, with the exception of the double heterozygous strain (centre), which was compared to all other genotypes. Note that the strain *erg5/ERG5 erg6/erg6* was later found to be homozygous for the mutation in *ERG5*, likely due to a loss of heterozygosity event. All underlying raw data and analyses can be found in Ono *et al.* (2016).

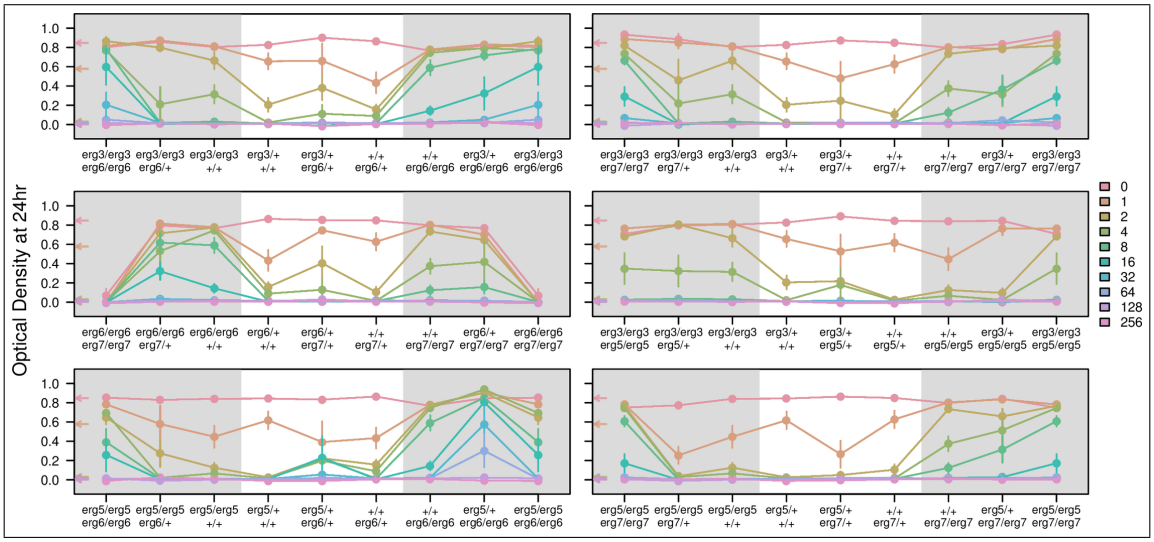

Figure B.5: Optical density after 24 hours of growth for diploid strains in a range of concentrations of nystatin. Colours go from red to purple, through blues, from lowest to highest concentrations of nystatin. Lines connect different mutants in the same concentration of nystatin. Mutant strains are ordered one mutational step apart along the x-axis, with the homozygous double mutant at both ends. Sections shaded in grey represent mutants carrying at least one homozygous mutation. Differences in OD between mutants were not tested statistically and are all represented by solid lines (in contrast to Fig 3.5). Arrows on the y-axes indicate the OD of the ancestral strain. All replicates were averaged, and error bars denote the standard error. Note that tolerance was assayed in the *erg5/erg5 erg6/erg6* homozygous double mutant before we determined that it was likely polymorphic; these points may thus be underestimates (see Table B.2 for details). Also note that the strain *erg5/ERG5 erg6/erg6* was later found to be homozygous for the mutation in *ERG5*, likely due to a loss of heterozygosity event. All underlying raw data and analyses can be found in Ono *et al.* (2016).

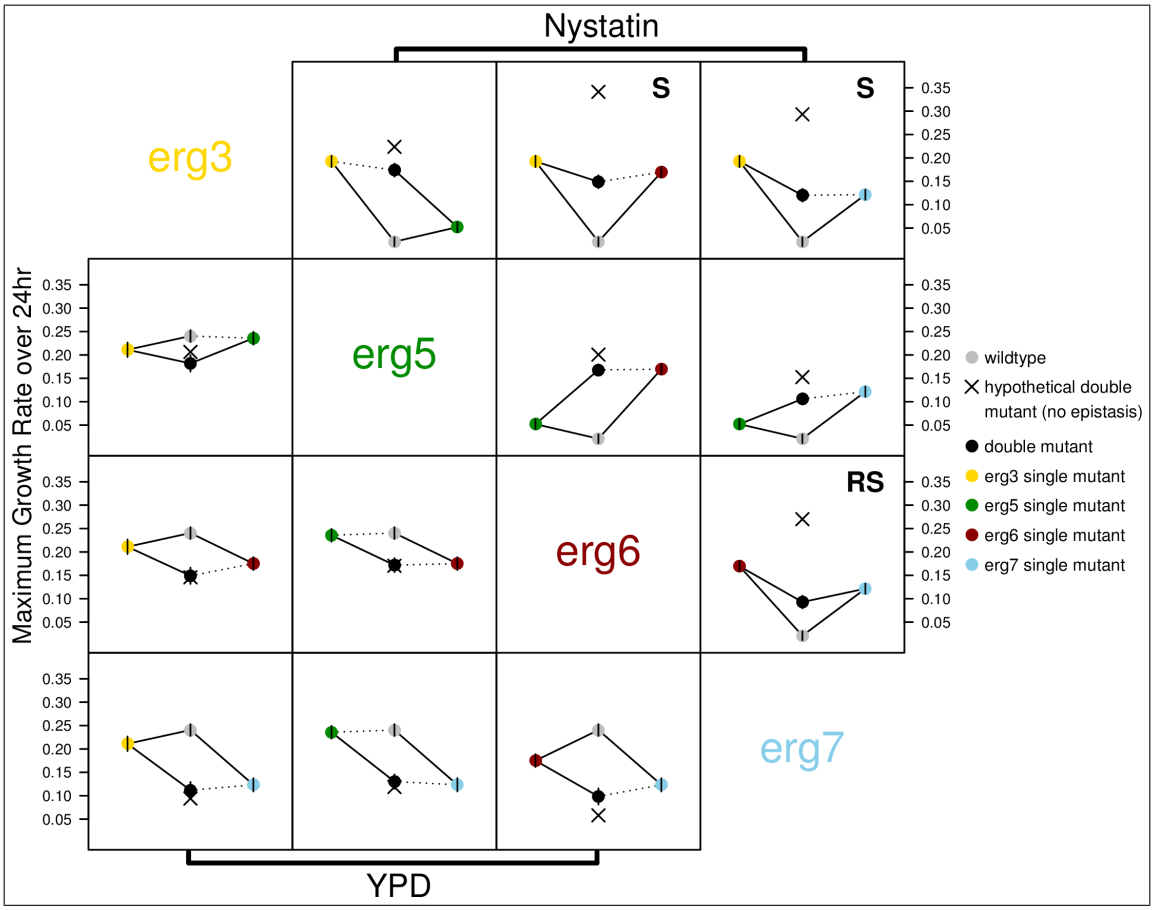

Figure B.6: Maximum growth rate of haploid strains in nystatin2 (above diagonal) and YPD (below diagonal) when including outliers. Points are the fitted least-squares means of the maximum growth rates, determined in the mixed-effects model. ×'s denote the additive fitness null expectation for the double mutant, i.e., with no epistasis. Each single mutant is coloured differently, the double mutant is black, and the ancestor is grey. Vertical bars represent 95% confidence intervals of the fitted least-squares mean. Solid lines indicate significant comparisons, while dotted lines are non-significant comparisons. Combinations showing significant sign (S) and reciprocal sign (RS) epistasis are indicated by the presence of the abbreviation at the top of the panel. All underlying raw data and analyses can be found in Ono *et al.* (2016).

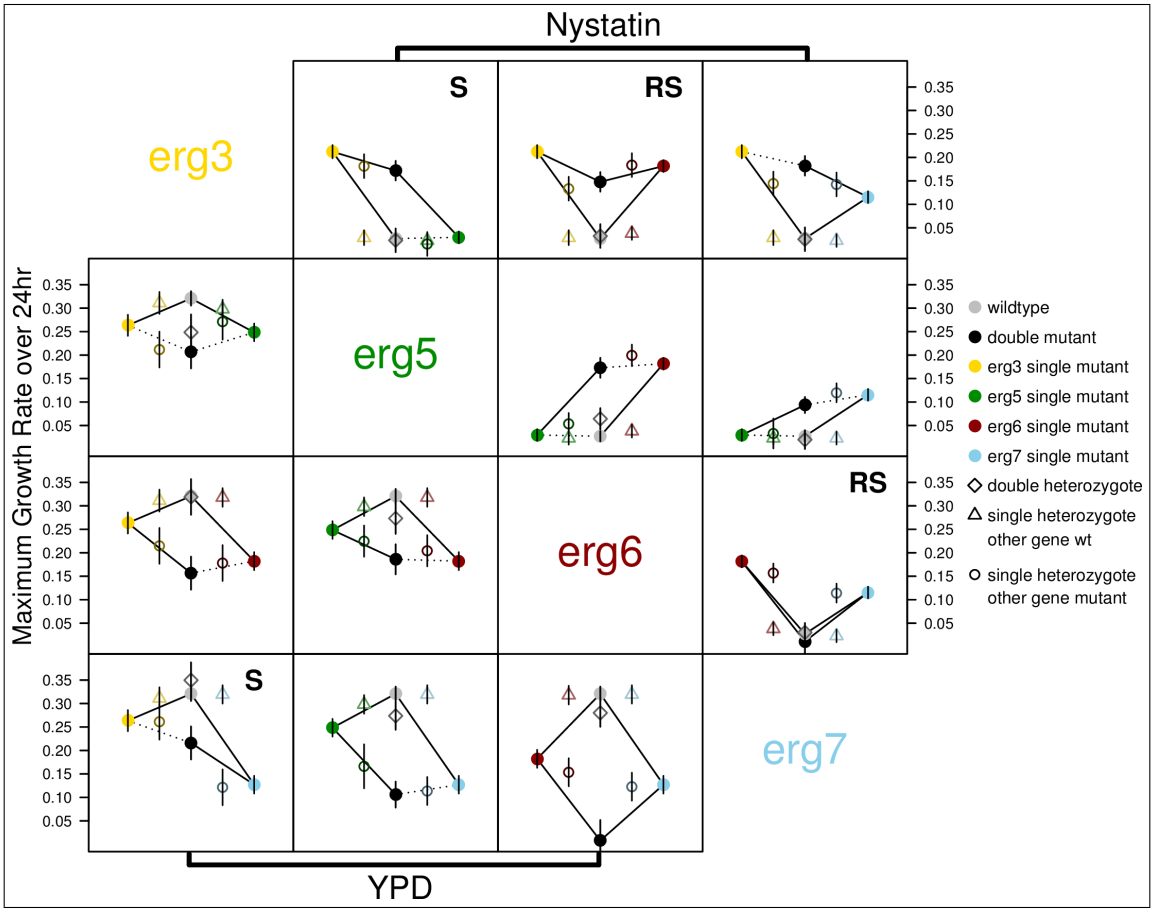

Figure B.7: Maximum growth rate of diploid strains in nystatin2 (above diagonal) and YPD (below diagonal) when including outliers. Points are the fitted least-squares means of the maximum growth rates, with closed circles determined in the mixed-effects model including only homozygous strains and open symbols from the model that includes heterozygous strains (open diamonds: double heterozygotes; open triangles: single heterozygotes that are wildtype at the other gene; open circles: single heterozygotes that are homozygous mutants at the other gene). Points and bars are otherwise as in Fig 3.3 and Fig B.6. All symbols are coloured intermediately according to genotype and arrayed along the x-axis so as to lie between the two strains that are genotypically most similar to it. Solid lines indicate significant comparisons in tests run including only homozygous strains while dotted lines are non-significant comparisons. See Fig 3.3 or Fig B.6 for further graphical details. Note that the strain *erg5/ERG5 erg6/erg6* was later found to be homozygous for the mutation in *ERG5*, likely due to a loss of heterozygosity event. All underlying raw data and analyses can be found in Ono *et al.* (2016).

Appendix C

Appendix for Chapter 4: The limit to evolutionary rescue depends on ploidy in yeast exposed to nystatin

C.1 Strain Differences

Despite not being judged as putative mutants, many diploid wells did show growth in the initial acquisition experiments. There was a significant association between the identity of a strain and whether or not it grew in the acquisition experiments (χ^2 contingency test using *chisq.test* in the R package *stats* [R CORE TEAM 2016]: Acquisition Experiment 1: χ^2 = 61.58, df = 2, p-value = 4.26×10^{-14}; Acquisition Experiment 2: χ^2 = 72.7, df = 2, p-value < 10^{-15}; Acquisition Experiment 3: χ^2 = 168.38, df = 2, p-value < 10^{-15}). In the first two acquisition experiments (performed in the BY strains and YPDnystatin4), *MAT*α grew the most, proportionally, followed by *MATa* and the diploids, which had similar growth. In Acquisition Experiment 3 (performed in the W303 strains and SCnystatin4), *MATa* grew the most followed by the diploids and both grew much more than the *MAT*α wells. This difference in growth was observed despite the additional opportunity given to 80 of the *MAT*α wells that were sampled and run in duplicate in the acquisition experiment. The difference in growth of the *MAT*α wells was the main observed inconsistency between the two genetic backgrounds used (BY vs. W303), and can be accounted for by the respiratory-deficiency of our copy of the MJM36 strain. The associations remained significant when we grouped all strains of a single ploidy, with haploids growing more often than diploids in the first two acquisition experiments but not the third (because of the poor growth of *MAT*α).

However, there was a difference in the distribution of days until growth between strains (χ^2 contingency test using 'chisq.test' with a simulated p-value based on 10,000 replicates: Acquisition Experiment 1: χ^2 = 49.82, p-value < 1.00×10^{-4}; Acquisition Experiment 2: χ^2 = 90.69, p-value < 1.00×10^{-4}; Acquisition Experiment 3: χ^2 = 159.86, p-value < 1.00×10^{-4}) (Fig. C.2), and these associations remained significant when we grouped all strains of a single ploidy. *MATa* populations had the lowest mean number of days until growth, followed by *MAT*α and then the diploids (details of comparisons in Table C.1). In the follow-up growth assays, OD72 (and therefore whether a population was judged to be putatively nystatin-resistant) was correlated with number of days until growth, decreasing with increasingly later day of acquisition (Kendall's rank correlation using *cor.test* in the R package *stats* [R CORE TEAM 2016], Experiment 1: τ = -0.29, z = -5.90, p-value = 3.62×10^{-9}; Experiment 2: τ = -0.45, z = -7.80, p-value = 6.04×10^{-15}; Experiment 3: τ = -0.46, z = -9.52, p-value < 10^{-15}, Fig C.2). We conclude that later-growing strains are less likely to be

true mutants. These wells may instead be growing due to the degradation or inactivation of nystatin in the medium over time (see Section 4.3.2).

C.2 Mutant Coverage

The following Mathematica package was used to carry out the calculations and is available upon request.

Diploidy limits evolutionary rescue in yeast exposed to nystatin

Supplementary *Mathematica* file for Ono et al.

Chance that a mutation was sampled across the acquisition experiments

Deep Well Boxes

Here we model the growth of a population from a single cell established on an agar plate (rich medium), picked as a colony, and grown to saturation in 150uL rich medium, where cell densities were estimated by hemacytometer to be ~7.03×10^7cells/mL for the BY "source" population (~7.42×10^7cells/mL for the W303 "source" population), from which a sample of 10uL is taken to establish the population in each deep well.

We then calculate the expected number of mutations that would occur within at least one diploid cell within each deep well and across the entire experiment. To be conservative, we use the lower reported mutation rate per basebair of 1.67×10^{-10} from Zhu et al. (2014, PNAS), rather than the higher 3.3×10^{-10} from Lynch et al. (2008, PNAS).

Parameters:

μ = mutation rate per basepair per cell division
N1 = population size at saturation in 150uL rich medium for growth in bioscreen
f = fraction of population sampled to found a lineage (0.0667 = 10 ul/150 uL)
L = # diploid deep wells (619: {191-13,286,155} in the three acquisition experiments in deep well boxes)
ORF = average length of an ORF (1385 from Hurowitz, E. H., & Brown, P. O. 2003 Genome Biology)

```
tryμZHU = 1.67 × 10^-10;
tryμLYNCH = 3.3 × 10^-10;
tryNBY = 7.03 * 10^7 * 0.15;
(*Estimated population density per mL for BY and scaling to 150uL YPD.*)
tryN303 = 7.42 * 10^7 * 0.15;
(*Estimated population density per mL for W303 and scaling to 150uL SC.*)
tryf = 0.0667;
tryL = {191 - 13, 286, 155}; (*Number of diploid wells per acquisition experiment,
excluding ones later found to have been haploid contaminated*)
tryORF = 1385;
```

Number of cell cycles required to produce source population of N1 cells:

```
cycles = Log[2, N1]
```

$$\frac{\text{Log}[N1]}{\text{Log}[2]}$$

In cases where we are performing numerical sums and require an integer number, we round down the number of cycles (rounding down is slightly conservative):

```
tryc = Floor[cycles /. N1 → tryNBY]
```

```
23
```

This is the same integer number of cycles for the W303 strain:

```
Floor[cycles /. N1 → tryN303]
```

```
23
```

Total number of cell divisions involved (1 cell division from 1 → 2 cells, 2 cell divisions from 2 → 4 cells, etc):

```
divisions = Sum[2^i, {i, 0, cycles - 1}]
```

$-1 + N1$

For example, to go from 1 → 4 cells involves a total of 3 divisions (→ 8 cells would involve 7 dividing cells: one 1→2, two 2→4, and four 4→8):

```
divisions /. N1 → {4, 8}
```

{3, 7}

For the first two acquisition experiments:

```
divisions /. N1 → tryNBY // N
```

1.0545×10^{7}

The chance that NONE of these cell divisions produced a mutation at a particular site in a diploid (bearing 2μ mutations per cell division across the two homologues) is:

```
(1 - 2 μ)^divisions;
```

Given that μ is small, this is very nearly:

```
nomutantsinsource[μ_, N1_] = ℯ^(-2 μ*N1);
```

Using a per-basepair mutation rate of 1.67×10^{-10} (Zhu et al. 2014), the chance of no mutations at a single site within the source pool for a single well would be:

```
nomutantsinsource[tryμZHU, tryNBY]
```

0.996484

The distribution of mutant cell numbers in the source population is broad and very skewed (a "jackpot distribution"), and it is possible that the mutation hit early and generated many mutant cells. To account for this mutational distribution, the probability that a mutation at one specific site occurs in the jth cell cycle (going from $m = 2^{j-1}$ cells to 2^j cells) is:

```
prob[j_, μ_] = 1 - (1 - 2 μ)^m /. m → 2^(j-1);
```

based on one minus the probability that no mutation hits. (Technically, this allows for the possibility that more than one hit would occur at the exact same site in different cells in the same cell cycle, but the chance is exceedingly unlikely.)

If a mutation does occur in the jth cycle (i.e., among the 2^j cells that result in this cycle, where one is a new mutant), the fraction of the source population that will be mutant (assumed to be unaffected by selection prior to placement in nystatin) is:

```
frac[j_] = 1 / 2^j;
```

Thus, prob[j,μ] gives us the probability distribution for the fraction, frac[j], of mutant cells in the source population (amounting to a number of mutant cells $N1 / 2^j$), with the probability of at least one mutant cell at a particular site equalling:

```
mutantprob[μ_, N1_] = Sum[prob[j, μ], {j, 1, cycles}]
```

$$\sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(1 - (1 - 2\mu)^{2^{-1+j}}\right)$$

For the mutation rate of Zhu et al. and the population size estimated for BY, the probability of at least one mutant cell in the source population is:

```
mutantprob[tryμZHU, tryNBY]
```

```
0.00280049
```

The expected number of mutant cells is the chance of a mutation occurring at cycle j and the number of cells that result (N1*frac[j]), summed over all cell cycles:

```
Sum[prob[j, tryμZHU] * tryNBY * frac[j], {j, 1, tryc}]
```

```
0.0405009
```

Finally, we calculate the probability that one or more mutant cells will be placed in a well of a deep well box, given that a fraction, f, of the N1 cells were sampled:

```
probhit[μ_, N1_, f_, c_] =
 1 - (1 - mutantprob[μ, N1]) - Sum[prob[j, μ] * (1 - frac[j])^f*N1 , {j, 1, c}]
(*We calculate the probability of sampling some mutant
 cells as one minus the probability of sampling none,
either because no mutations occur (1-mutantprob[μ,N1]) or because
 mutations occur in cycle c but are not sampled (the sum)*)
```

$$\sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right) - \sum_{j=1}^{c} \left(1 - 2^{-j}\right)^{f\,N1} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right)$$

```
probhit[tryμZHU, tryNBY, tryf, tryc]
```

```
0.000555156
```

Given L independent deep wells (each started from a different colony), where each lineage is started with a fraction, f, of its own source population, the probability that at least one mutant cell will be sampled into at least one of the deep well populations would be:

```
1 - (1 - probhit[μ, N1, f, c])^L
```

$$1 - \left(1 - \sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right) + \sum_{j=1}^{c} \left(1 - 2^{-j}\right)^{f\,N1} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right)\right)^{L}$$

```
1 - (1 - probhit[tryμZHU, tryNBY, tryf, tryc])^Total[tryL]
```

```
0.290885
```

The expected number of deep wells with at least one mutant cell bearing a mutation at the focal site is then:

```
Total[tryL] * probhit[tryμZHU, tryNBY, tryf, tryc]
 (*Expected number of wells with mutations.*)
```

```
0.343642
```

This is only slightly larger if we account for the higher estimated population size for W303:

```
Sum[tryL[[i]] * probhit[tryμZHU, tryNBY, tryf, tryc], {i, 1, 2}] +
 tryL[[3]] * probhit[tryμZHU, tryN303, tryf, tryc]
```

```
0.346963
```

But the estimate would almost double if the mutation rate were closer to that inferred by Lynch et al. (2008):

```
Total[tryL] * probhit[tryμLYNCH, tryNBY, tryf, tryc]
 (*Expected number of wells with mutations.*)
```

```
0.678879
```

Returning to our defaults, if we multiply by the average ORF length, we would expect 479 mutations per ORF to have arisen across all of the deep wells:

```
Total[tryL] * probhit[tryμZHU, tryNBY, tryf, tryc] * tryORF
```

```
475.944
```

Of these mutations per ORF, we expect each gene would have, on average, 23 premature stop codons somewhere in one of the wells (fraction of non-sense mutations in a gene estimated in Gerstein et al. (2015)'s Supplementary *Mathematica* package, available on DRYAD, DOI: https://doi.org/10.5061/dryad.5gp25):

```
Total[tryL] * probhit[tryμZHU, tryNBY, tryf, tryc] * tryORF * 0.048831111111111124`
```

```
23.2409
```

Similarly, Each gene would have, on average, 350 AA changes somewhere in one of the wells (fraction of non-synonymous mutations in a gene estimated in Gerstein et al. (2015)'s Supplementary *Mathematica* package, available on DRYAD, DOI: https://doi.org/10.5061/dryad.5gp25):

```
Total[tryL] * probhit[tryμZHU, tryNBY, tryf, tryc] * tryORF * 0.7363755555555553`
```

```
350.473
```

■ Flasks

Here we model the growth of a population from a single cell established on an agar plate (YPD), picked as a colony, and grown to saturation in 10mL YPD, which corresponds to a population size of $\sim 7 \times 10^8$ cells ("source" population), from which a sample of 1mL is taken to establish each individual lineage (the "founding" population).

Other parameters as above.

```
tryμZHU = 1.67 × 10^-10;
tryNBY = 7.03 * 10^7 * 10;
(*Using estimated population density per mL and scaling up to 10mL*)
tryf = 0.1; (*1mL into 10mL*)
tryL = 10;(*Number of diploid flasks*)
tryORF = 1385;
```

Number of cell cycles required to produce source population of N1 cells:

```
cycles = Log[2, N1]
```

$$\frac{\text{Log}[N1]}{\text{Log}[2]}$$

In cases where we are performing numerical sums and require an integer number, we round down the number of cycles (rounding down is slightly conservative):

```
tryc = Floor[cycles /. N1 → tryNBY]
```

```
29
```

Total number of cell divisions involved (1 cell division from 1 -> 2 cells, 2 cell divisions from 2 -> 4 cells, etc):

```
divisions = Sum[2^i, {i, 0, cycles - 1}]
```

```
-1 + N1
```

The chance that NONE of these cell divisions produced a mutation at a particular site in a diploid (bearing 2μ mutations per cell division across the two homologues) is:

```
(1 - 2 μ)^divisions;
```

The distribution of mutant cell numbers in the source population is broad and very skewed (a "jackpot distribution"), and it is possible that the mutation hit early and generated a lot of mutant cells. To account for this mutational distribution, the probability that a mutation at one specific site occurs in the jth cell cycle (going from $m = 2^{j-1}$ cells to 2^j cells) is:

```
prob[j_, μ_] = 1 - (1 - 2 μ)^m /. m → 2^(j-1);
```

based on one minus the probability that no mutation hits. (Technically, this allows for the possibility that more than one hit would occur at the exact same site in different cells in the same cell cycle, but the chance is exceedingly unlikely.)

If a mutation does occur in the jth cycle (i.e., among the 2^j cells that result in this cycle, where one is a new mutant), the fraction of the source population that will be mutant (assumed to be unaffected by selection prior to placement in nystatin) is:

```
frac[j_] = 1 / 2^j;
```

Thus, prob[j,μ] gives us the probability distribution for the fraction, frac[j], of mutant cells in the source population (amounting to a number of mutant cells $N1 / 2^j$), with the probability of at least one mutant cell at a particular site equalling:

```
mutantprob[μ_, N1_] = Sum[prob[j, μ], {j, 1, cycles}]
```

$$\sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right)$$

For the mutation rate of Zhu et al. and the population size estimated for BY, the probability of at least one mutant cell in the source population is:

```
mutantprob[tryμZHU, tryNBY]
```

```
0.17409
```

The expected number of mutant cells is the chance of a mutation occurring at cycle j and the number of cells that result (N1*frac[j]), summed over all cell divisions:

```
Sum[prob[j, tryμZHU] * tryNBY * frac[j], {j, 1, tryc}]
```

```
3.39431
```

Finally, we calculate the probability that one or more mutant cells will be placed in the flask, given that a fraction, f, of the N1 cells were sampled:

```
probhit[μ_, N1_, f_, c_] =
 1 - (1 - mutantprob[μ, N1]) - Sum[prob[j, μ] * (1 - frac[j])^(f*N1) , {j, 1, c}]
(*We calculate the probability of sampling some mutant
 cells as one minus the probability of sampling none,
either because no mutations occur (1-mutantprob[μ,N1]) or because
 mutations occur in cycle c but are not sampled (the sum)*)
```

$$\sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right) - \sum_{j=1}^{c} \left(1 - 2^{-j}\right)^{f\,N1} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right)$$

```
probhit[tryμZHU, tryNBY, tryf, tryc]
```

```
0.0473538
```

Given L independent flasks (each started from a different colony), where each lineage is started with a fraction, f, of its own source population, the probability that at least one mutant cell will be sampled into at least one of the flask populations would be:

```
1 - (1 - probhit[μ, N1, f, c])^L
```

$$1 - \left(1 - \sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right) + \sum_{j=1}^{c} \left(1 - 2^{-j}\right)^{f\,N1} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right)\right)^{L}$$

```
1 - (1 - probhit[tryμZHU, tryNBY, tryf, tryc])^tryL
```

```
0.384375
```

The expected number of flasks with at least one mutant cells bearing a mutation at the focal site is then:

```
tryL * probhit[tryμZHU, tryNBY, tryf, tryc](*Expected number of flasks with mutations.*)
```

```
0.473538
```

If we multiply by the average ORF length, we would expect 656 mutations per ORF to have arisen across the 10 flasks:

```
tryL * probhit[tryμZHU, tryNBY, tryf, tryc] * tryORF
```

```
655.851
```

Of these mutations per ORF, we expect each gene would have, on average, 32 premature stop codons somewhere in one of the flasks (fraction of non-sense mutations in a gene estimated in Gerstein et al. (2015)'s Supplementary *Mathematica* package, available on DRYAD, DOI: https://doi.org/10.5061/dryad.5gp25):

```
tryL * probhit[tryμZHU, tryNBY, tryf, tryc] * tryORF * 0.048831111111111124`
```

```
32.0259
```

Similarly, Each gene would have, on average, 483 AA changes somewhere in one of the flasks (fraction of non-synonymous mutations in a gene estimated in Gerstein et al. (2015)'s Supplementary *Mathematica* package, available on DRYAD, DOI: https://doi.org/10.5061/dryad.5gp25):

```
tryL * probhit[tryμZHU, tryNBY, tryf, tryc] * tryORF * 0.7363755555555553`
```

```
482.952
```

Conclusion

The above calculations inform us that each site within the genome is likely to have been hit by a nucleotide changing mutation (expected number of deep well hits = 0.35; expected number of flask hits = 0.47; expected number of total hits = 0.82). As a consequence, we are likely to have sampled ~888 non-synonymous or non-sense mutations for each gene within the genome over the course of the experiment.

Caveats: Of course, the fact that a single cell of the right genotype is sampled into one of the wells doesn't mean that it will necessarily establish; it may die before dividing. Plus, the above calculations used the average mutation rate; sites with lower mutation rates are less likely to have been sampled. Furthermore, we only calculate the chance that a site mutates, not the chance of having sampled all three possible alternative nucleotides (which would require accounting for differences in transition and transversion mutation rates). Finally, the above calculations are only for SNPs and ignore more complex mutations (indels, rearrangements, LOH, etc.)

Nevertheless, the above calculations demonstrate that the acquisition experiments had a reasonable chance of exploring SNP mutations at most sites within the genome.

Chance that a two-step mutation would arise

Deep Well Boxes

Here we modify the above to calculate the chance of observing two mutations within the same gene within a well, providing resistance even if only recessive mutations are available. We consider two cases, either where the secondary mutation can occur anywhere within the same ORF (in the homologue) or where the secondary mutation is a loss-of-heterozygosity event (estimated to occur by mitotic recombination at a rate of ~0.8×10^{-4}; Mandegar & Otto, 2007, Proc Roy Soc B; only half of which is assumed to lead to the homozygous recessive mutant).

Additional parameters:

$\mu 2$ = secondary mutation rate (assumed to be either $\mu \times$ ORF or $0.8 \times 10^{-4} / 2$)

N1 = population size at saturation in 150uL rich medium for growth in bioscreen
f = fraction of population sampled to found a lineage (0.0667 = 10 ul/150 uL)
L = # diploid deep wells (619: {191-13,286,155} in the three acquisition experiments in deep well boxes)
ORF = average length of an ORF (1385 from Hurowitz, E. H., & Brown, P. O. 2003 Genome biology, 5(1), R2.)

```
tryμZHU = 1.67 × 10^-10;
tryμLYNCH = 3.3 × 10^-10;
tryNBY = 7.03 * 10^7 * 0.15;
(*Estimated population density per mL for BY and scaling to 150uL YPD.*)
tryN303 = 7.42 * 10^7 * 0.15;
(*Estimated population density per mL for W303 and scaling to 150uL SC.*)
tryf = 0.0667;
tryL = {191 - 13, 286, 155}; (*Number of diploid wells per acquisition experiment,
excluding ones later found to have been haploid contaminated*)
tryORF = 1385;
tryμ2a = tryμZHU * tryORF;
(*Assuming that any other mutation in the same gene would inactivate,
a conservative assumption.*)
tryμ2b = 0.8 × 10^-4 / 2; (*Assuming a secondary LOH event,
half of which lead to the homozygous recessive mutant.*)
```

Number of cell cycles required to produce source population of N1 cells:

```
cycles = Log[2, N1]
```

$$\frac{\text{Log}[N1]}{\text{Log}[2]}$$

In cases where we are performing numerical sums and require an integer number, we round down the number of cycles (rounding down is slightly conservative):

```
tryc = Floor[cycles /. N1 → tryNBY]
```

```
23
```

Total number of cell divisions involved (1 cell division from 1 → 2 cells, 2 cell divisions from 2 → 4 cells, etc):

```
divisions = Sum[2^i, {i, 0, cycles - 1}]
```

```
-1 + N1
```

The chance that NONE of these cell divisions produced a mutation at a particular site in a diploid (bearing 2μ mutations per cell division across the two homologues) is:

```
(1 - 2 μ)^divisions;
```

The distribution of mutant cell numbers in the source population is broad and very skewed (a "jackpot distribution"), and it is possible that the mutation hit early and generated a lot of mutant cells. To account for this mutational distribution, the probability that a mutation at one specific site occurs in the jth cell cycle (going from $m = 2^{j-1}$ cells to 2^j cells) is:

```
prob[j_, μ_] = 1 - (1 - 2 μ)^m /. m → 2^(j-1);
```

based on one minus the probability that no mutation hits. (Technically, this allows for the possibility that more than one hit would occur at the exact same site in different cells in the same cell cycle, but the chance is exceedingly unlikely.)

If a mutation does occur in the jth cycle (i.e., among the 2^j cells that result in this cycle, where one is a new mutant), the fraction of the source population that will be mutant (assumed to be unaffected by selection prior to placement in nystatin) is:

```
frac[j_] = 1 / 2^j;
```

The number of remaining cell cycles is "cycles-j", during which there will be $2^{(\text{cycles-j})} - 1$ divisions among the cells that already carry the first mutation (using the same logic used above to get "divisions"), so that the chance that a mutation occurs and then bears a secondary mutation is:

```
Sum[prob[j, μ] * μ2 * (2^(cycles-j) - 1), {j, 1, cycles}]
```

$$\sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(-1 + 2^{-j}\, N1\right) \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right) \mu 2$$

For a secondary mutation in the same ORF:

```
Sum[prob[j, tryμZHU] * tryμ2a * (2^(tryc-j) - 1), {j, 1, tryc}]
```

$$6.80428 \times 10^{-9}$$

For a secondary LOH event:

```
Sum[prob[j, tryμZHU] * tryμ2b * (2^(tryc-j) - 1), {j, 1, tryc}]
```

$$1.17673 \times 10^{-6}$$

The above just calculates the chance that a two-step mutation is in the source population. The chance that it will be sampled requires that we first calculate the expected number of two-step mutant cells in the source population.

The chance that a first mutation occurs in the jth cell cycle (going from $m = 2^{j-1}$ cells to 2^j cells) and then a secondary mutation occurs in the kth cell cycle among the cells that bear the first mutation (going from $n = 2^{k-j-1}$ mutant cells to 2^{k-j} mutant cells) is:

```
prob2[j_, k_, μ_, μ2_] = (1 - (1 - 2 μ)^m) (1 - (1 - μ2)^n) /. m → 2^(j-1) /. n → 2^(k-j-1);
```

(this assumes that the chance of both mutations happening in the same cell division is negligible and assumes that only mutations in the homologue can generate resistance).

If a secondary mutation does occur in the kth cycle after the first one in the jth cycle (when there are 2^k cells), the fraction of the source population that will be mutant (assumed to be unaffected by selection prior to placement in nystatin) is:

```
frac2[j_, k_] = 1 / 2^k;
```

Thus, prob2 gives us the probability distribution for the fraction, frac2, of secondary mutant cells in the source population (amounting to a number of mutant cells $N1 / 2^k$), with the probability of at least one two-step mutant cell at a particular site equalling:

```
mutantprob[μ_, μ2_, N1_] = Sum[prob2[j, k, μ, μ2], {j, 1, cycles - 1}, {k, j + 1, cycles}]
```

$$-1 + \sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \sum_{k=1+j}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right) \left(1 - (1 - \mu 2)^{2^{-1-j+k}}\right)$$

For a secondary mutation in the same ORF:

```
mutantprob[tryμZHU, tryμ2a, tryNBY]
```

$$6.70832 \times 10^{-9}$$

For a secondary LOH event:

```
mutantprob[tryμZHU, tryμ2b, tryNBY]
```

$$8.77605 \times 10^{-7}$$

Finally, we calculate the probability that one or more secondary mutant cells will be placed in the deep well box, given that a fraction, f, of the N1 cells were sampled:

```
probhit2[μ_, μ2_, N1_, f_, c_] = 1 - (1 - mutantprob[μ, μ2, N1]) -
   Sum[prob2[j, k, μ, μ2] * (1 - frac2[j, k])^(f*N1) , {j, 1, c - 1}, {k, j + 1, c}]
(*We calculate the probability of sampling some two-
 step mutant cells as one minus the probability of sampling none,
either because no two-step mutations occur (1-mutantprob[μ,μ2,N1]) or because two-
 step mutations occur but are not sampled (the sum)*)
```

$$-1+\sum_{j=1}^{\frac{\text{Log[N1]}}{\text{Log[2]}}}\ \sum_{k=1+j}^{\frac{\text{Log[N1]}}{\text{Log[2]}}}\left(1-(1-2\,\mu)^{2^{-1+j}}\right)\left(1-(1-\mu 2)^{2^{-1-j+k}}\right)-$$

$$\sum_{j=1}^{-1+c}\ \sum_{k=1+j}^{c}\left(1-2^{-k}\right)^{f\,N1}\left(1-(1-2\,\mu)^{2^{-1+j}}\right)\left(1-(1-\mu 2)^{2^{-1-j+k}}\right)$$

For a secondary mutation in the same ORF:

```
probhit2[tryμZHU, tryμ2a, tryNBY, tryf, tryc]
```

$$1.24781\times 10^{-9}$$

For a secondary LOH event:

```
probhit2[tryμZHU, tryμ2b, tryNBY, tryf, tryc]
```

$$1.72446\times 10^{-7}$$

Given L independent deep wells (each started from a different colony), where each lineage is started with a fraction, f, of its own source population, the probability that at least one secondary mutant cell will be sampled into at least one of the deep well populations would be:

```
1 - (1 - probhit2[μ, μ2, N1, f, c])^L
```

$$1-\left(1-\sum_{j=1}^{-1+\frac{\text{Log[N1]}}{\text{Log[2]}}}\ \sum_{k=1+j}^{\frac{\text{Log[N1]}}{\text{Log[2]}}}\left(1-(1-2\,\mu)^{2^{-1+j}}\right)\left(1-(1-\mu 2)^{2^{-1-j+k}}\right)+\sum_{j=1}^{-1+c}\ \sum_{k=1+j}^{c}\left(1-2^{-k}\right)^{f\,N1}\left(1-(1-2\,\mu)^{2^{-1+j}}\right)\left(1-(1-\mu 2)^{2^{-1-j+k}}\right)\right)^{L}$$

For a secondary mutation in the same ORF:

```
1 - (1 - probhit2[tryμZHU, tryμ2a, tryNBY, tryf, tryc])^Total[tryL]
```

$$7.72393\times 10^{-7}$$

For a secondary LOH event:

```
1 - (1 - probhit2[tryμZHU, tryμ2b, tryNBY, tryf, tryc])^Total[tryL]
```

```
0.000106739
```

■ Flasks

Here we modify the above to calculate the chance of observing two mutations within the same gene within a flask, providing resistance even if only recessive mutations are available. We consider two cases, either where the secondary mutation can occur anywhere within the same ORF or where the secondary mutation is a loss-of-heterozygosity event (estimated to occur by mitotic recombination at a rate of $\sim 0.8\times 10^{-4}$; Mandegar & Otto, 2007, Proc Roy Soc B; only half of which is assumed to lead to the homozygous recessive mutant).

Additional parameters:

$\mu 2$ = secondary mutation rate (assumed to be either $\mu \times$ ORF or $0.8\times 10^{-4}/2$)

N1 = population size at saturation in 10mL YPD

f = fraction of population sampled to found a lineage (0.1 = 1 mL/10 mL)

L = # diploid flasks in expt (10)

ORF = average length of an ORF (1385 from Hurowitz, E. H., & Brown, P. O. 2003 Genome biology, 5(1), R2.)

```
tryμZHU = 1.67 × 10^-10;
tryNBY = 7.03 * 10^7 * 10;
(*Using estimated population density per mL and scaling up to 10mL*)
tryf = 0.1; (*1mL into 10mL*)
tryL = 10;
tryORF = 1385;
```

```
tryμ2a = tryμZHU * tryORF;
(*Assuming that any other mutation in the same gene would inactivate,
a conservative assumption.*)
tryμ2b = 0.8 × 10^-4 / 2; (*Assuming a secondary LOH event,
half of which lead to the homozygous recessive mutant.*)
```

Number of cell cycles required to produce source population of N1 cells:

```
cycles = Log[2, N1]
```

$$\frac{\text{Log}[N1]}{\text{Log}[2]}$$

In cases where we are performing numerical sums and require an integer number, we round down the number of cycles (rounding down is slightly conservative):

```
tryc = Floor[cycles /. N1 → tryNBY]
```

```
29
```

Total number of cell divisions involved (1 cell division from 1 → 2 cells, 2 cell divisions from 2 → 4 cells, etc):

```
divisions = Sum[2^i, {i, 0, cycles - 1}]
```

```
-1 + N1
```

The chance that NONE of these cell divisions produced a mutation at a particular site in a diploid (bearing 2μ mutations per cell division across the two homologues) is:

```
(1 - 2 μ)^divisions;
```

The distribution of mutant cell numbers in the source population is broad and very skewed (a "jackpot distribution"), and it is possible that the mutation hit early and generated a lot of mutant cells. To account for this mutational distribution, the probability that a mutation at one specific site occurs in the jth cell cycle (going from $m = 2^{j-1}$ cells to 2^j cells) is:

```
prob[j_, μ_] = 1 - (1 - 2 μ)^m /. m → 2^(j-1);
```

based on one minus the probability that no mutation hits. (Technically, this allows for the possibility that more than one hit would occur at the exact same site in different cells in the same cell cycle, but the chance is exceedingly unlikely.)

If a mutation does occur in the jth cycle (i.e., among the 2^j cells that result in this cycle, where one is a new mutant), the fraction of the source population that will be mutant (assumed to be unaffected by selection prior to placement in nystatin) is:

```
frac[j_] = 1 / 2^j;
```

The number of remaining cell cycles is "cycles-j", during which there will be $2^{(\text{cycles}-j)} - 1$ divisions among the cells that already carry the first mutation (using the same logic used above to get "divisions"), so that the chance that a mutation occurs and then bears a secondary mutation is:

```
Sum[prob[j, μ] * μ2 * (2^(cycles-j) - 1), {j, 1, cycles}]
```

$$\sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(-1 + 2^{-j}\, N1\right) \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right) \mu 2$$

For a secondary mutation in the same ORF:

```
Sum[prob[j, tryμZHU] * tryμ2a * (2^(tryc-j) - 1), {j, 1, tryc}]
```

$$5.59293 \times 10^{-7}$$

For a secondary LOH event:

```
Sum[prob[j, tryμZHU] * tryμ2b * (2^(tryc-j) - 1), {j, 1, tryc}]
```

```
0.0000967238
```

The above just calculates the chance that a two-step mutation is in the source population. The chance that it will be sampled requires that we first calculate the expected number of two-step mutant cells in the source population.

The chance that a first mutation occurs in the jth cell cycle (going from $m = 2^{j-1}$ cells to 2^j cells) and then a secondary mutation occurs in the kth cell cycle among the cells that bear the first mutation (going from $n = 2^{k-j-1}$ cells to 2^{k-j} mutant cells) is:

```
prob2[j_, k_, μ_, μ2_] = (1 - (1 - 2 μ)^m) (1 - (1 - μ2)^n) /. m → 2^(j-1) /. n → 2^(k-j-1);
```

(this assumes that the chance of both mutations happening in the same cell division is negligible and assumes that only mutations in the homologue can generate resistance).

If a secondary mutation does occur in the kth cycle after the first one in the jth cycle (when there are 2^k cells), the fraction of the source population that will be mutant (assumed to be unaffected by selection prior to placement in nystatin) is:

```
frac2[j_, k_] = 1 / 2^k;
```

Thus, prob2 gives us the probability distribution for the fraction, frac2, of secondary mutant cells in the source population (amounting to a number of mutant cells $N1 / 2^k$), with the probability of at least one two-step mutant cell at a particular site equalling:

```
mutantprob[μ_, μ2_, N1_] = Sum[prob2[j, k, μ, μ2], {j, 1, cycles - 1}, {k, j + 1, cycles}]
```

$$-1 + \sum_{j=1}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \sum_{k=1+j}^{\frac{\text{Log}[N1]}{\text{Log}[2]}} \left(1 - (1 - 2\,\mu)^{2^{-1+j}}\right) \left(1 - (1 - \mu 2)^{2^{-1-j+k}}\right)$$

For a secondary mutation in the same ORF:

```
mutantprob[tryμZHU, tryμ2a, tryNBY]
```

$$4.76763 \times 10^{-7}$$

For a secondary LOH event:

```
mutantprob[tryμZHU, tryμ2b, tryNBY]
```

```
0.0000562647
```

Finally, we calculate the probability that one or more secondary mutant cells will be placed in a flask, given that a fraction, f, of the N1 cells were sampled:

```
probhit2[μ_, μ2_, N1_, f_, c_] = 1 - (1 - mutantprob[μ, μ2, N1]) -
   Sum[prob2[j, k, μ, μ2] * (1 - frac2[j, k])^(f*N1) , {j, 1, c - 1}, {k, j + 1, c}]
(*We calculate the probability of sampling some two-
 step mutant cells as one minus the probability of sampling none,
either because no two-step mutations occur (1-mutantprob[μ,μ2,N1]) or because two-
 step mutations occur but are not sampled (the sum)*)
```

$$-1+\sum_{j=1}^{\frac{\text{Log[N1]}}{\text{Log[2]}}}\ \sum_{k=1+j}^{\frac{\text{Log[N1]}}{\text{Log[2]}}}\left(1-(1-2\,\mu)^{2^{-1+j}}\right)\left(1-(1-\mu 2)^{2^{-1-j+k}}\right)-$$

$$\sum_{j=1}^{-1+c}\ \sum_{k=1+j}^{c}\left(1-2^{-k}\right)^{f\,N1}\left(1-(1-2\,\mu)^{2^{-1+j}}\right)\left(1-(1-\mu 2)^{2^{-1-j+k}}\right)$$

For a secondary mutation in the same ORF:

```
probhit2[tryμZHU, tryμ2a, tryNBY, tryf, tryc]
```

1.27163×10^{-7}

For a secondary LOH event:

```
probhit2[tryμZHU, tryμ2b, tryNBY, tryf, tryc]
```

```
0.0000151389
```

Given L independent deep wells (each started from a different colony), where each lineage is started with a fraction, f, of its own source population, the probability that at least one secondary mutant cell will be sampled into at least one of the deep well populations would be:

```
1 - (1 - probhit2[μ, μ2, N1, f, c])^L
```

$$1-\left(1-\sum_{j=1}^{-1+\frac{\text{Log[N1]}}{\text{Log[2]}}}\ \sum_{k=1+j}^{\frac{\text{Log[N1]}}{\text{Log[2]}}}\left(1-(1-2\,\mu)^{2^{-1+j}}\right)\left(1-(1-\mu 2)^{2^{-1-j+k}}\right)+\sum_{j=1}^{-1+c}\ \sum_{k=1+j}^{c}\left(1-2^{-k}\right)^{f\,N1}\left(1-(1-2\,\mu)^{2^{-1+j}}\right)\left(1-(1-\mu 2)^{2^{-1-j+k}}\right)\right)^{L}$$

For a secondary mutation in the same ORF:

```
1 - (1 - probhit2[tryμZHU, tryμ2a, tryNBY, tryf, tryc])^tryL
```

1.27163×10^{-6}

For a secondary LOH event:

```
1 - (1 - probhit2[tryμZHU, tryμ2b, tryNBY, tryf, tryc])^tryL
```

```
0.000151379
```

■ Conclusion

Allowing secondary mutations to occur among the diploid cells that bear a first mutation, there is <0.0003 chance that a secondary mutant cell will be sampled in any of the deep wells or flasks in the experiment, assuming that the secondary mutation rate or LOH rate is $<\sim 10^{-4}$. We consider this to be a conservative estimate of the secondary mutation rate because we have assumed that any mutation in the same ORF on the homologue would generate resistance, as would mitotic recombination whenever it creates a homozygous mutant.

C.3 Supporting Table

Table C.1: Results of ANOVA and post-hoc Tukey tests comparing mean number of days until growth between the different strain types in each experiment. For the Tukey tests, estimates are reported with p-values in parentheses. Estimates are negative when the second group has a higher mean. Significant p-values are in bold.

	Acquisition Experiment 1	Acquisition Experiment 2	Acquisition Experiment 3
ANOVA			
F	14.45	61.25	97.55
df	2, 206	2, 161	2, 229
p-value	$\mathbf{1.35 \times 10^{-6}}$	$\mathbf{< 10^{-15}}$	$\mathbf{< 10^{-15}}$
Pairwise comparisons (Tukey)			
MATa - diploid	-2.26 (**< .0001**)	-2.86 (**< .0001**)	-1.87 (**< .0001**)
MATα - diploid	-1.13 (**0.0043**)	-2.54 (**< .0001**)	-0.27 (0.29)
MATa - *MAT*α	-1.13 (**0.023**)	-0.31 (0.75)	-1.60 (**< .0001**)

C.4 Supporting Figures

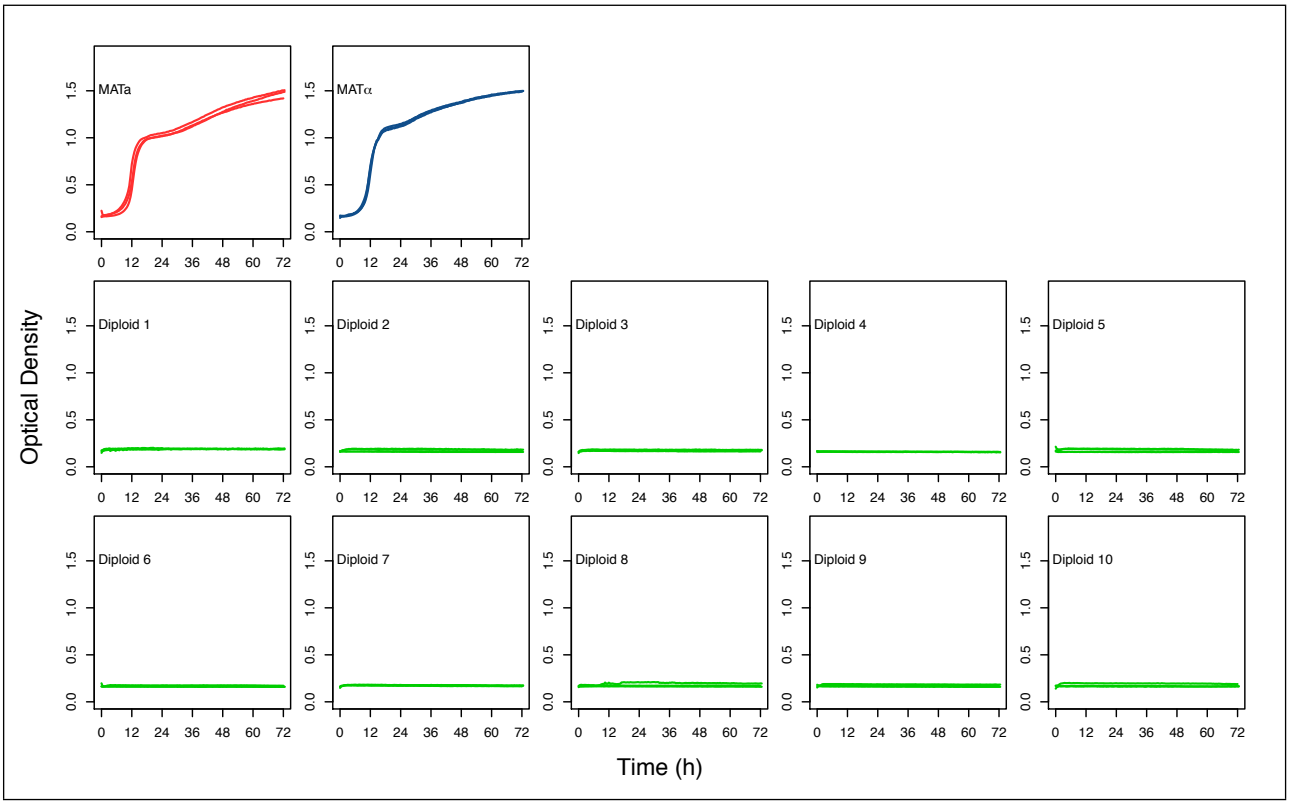

Figure C.1: Growth curves of all populations from larger volume flasks. All populations grew in the original acquisition experiment by Day 10, but only the haploid populations show reliable growth when re-tested in YPDnystatin4. No diploid populations grew over the 72 hour assay.

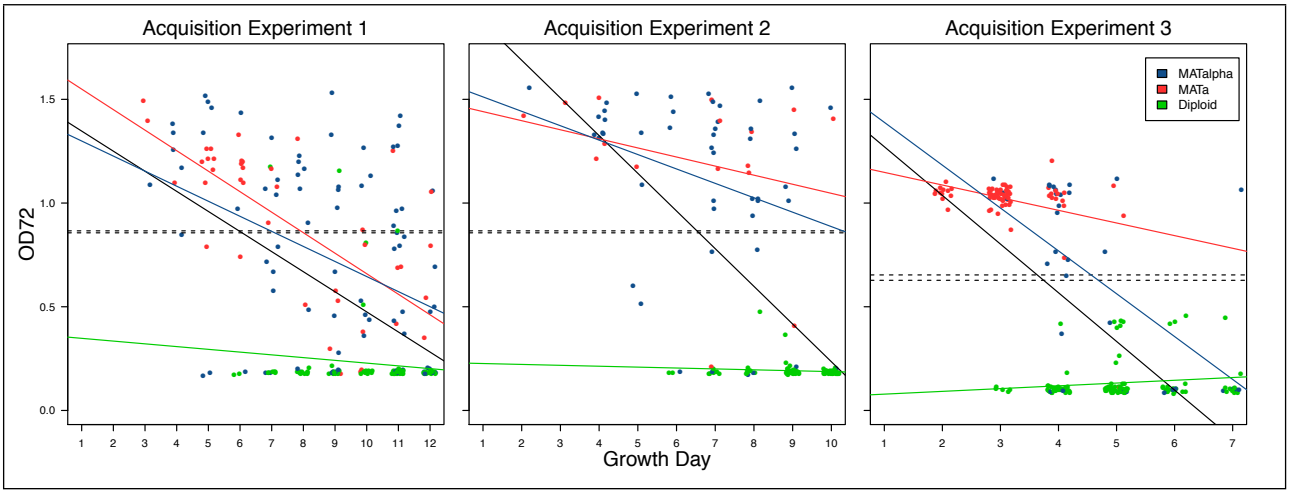

Figure C.2: Plot of mean OD72 when populations are re-tested in nystatin versus the number of days until growth was observed in the original acquisition experiments in deep well boxes. Black dashed lines indicate the range of cutoff values used for judging a potential mutant strain when growth was re-tested. Solid lines indicate the fitted regression lines from linear regressions (black: all populations considered together, coloured: model run using only the populations of that type). Note that the x-axis changes depending on the length of the acquisition experiment.