A characterization of adaptive mutations in yeast

by

Aleeza C. Gerstein

M.Sc. Zoology, University of British Columbia, 2006

B.Sc. Ecology & Evolution (Honours), The University of Western Ontario, 2003

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

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES
(Zoology)

The University Of British Columbia
(Vancouver)

August 2012

Abstract

Natural selection acts on phenotypes within populations, yet it is allele frequency changes at the genetic level that enable adaptation. To properly understand the evolutionary process we thus need to understand how the genotypic and external environments affect beneficial mutations and, in turn, affect the fitness of individuals. In this thesis I used the budding yeast, *Saccharomyces cerevisiae* to explore the genotypic basis, phenotypic diversity, and fitness effects of beneficial mutations in a variety of genotypic and external environments.

I first describe fitness experiments designed to elucidate the factor that allowed diploid mutants to overtake haploid populations during batch culture evolution. I compared haploid and diploid lines isolated at many time points using multiple growth phase and competitive fitness assays, yet diploids failed to demonstrate an advantage for any measure. I then conducted a related set of experiments that compared the rate of adaptation of haploid and diploid populations across seven different environments. I found that although haploid populations adapted faster than diploids in all environments, there was considerable variation between ploidy popualtions and among environments.

Experimental evolution results can be difficult to explain without knowledge of the specific mutations involved. The remainder of this thesis thus focused on a set of 20 unique beneficial mutations I acquired that confer tolerance to nystatin, a fungicide. The mutations are in four different genes that act close together late in the ergosterol biosynthesis pathway. Although the genetic basis of adaptation was narrow, lines that carried mutations in different genes were not equally tolerant to nystatin and were found to exhibit different gene-by-environment interactions. Surprisingly, the mutations had a larger effect size in nystatin in a haploid background than in a homozygous diploid background. I then show that the dominance of these mutations (i.e., the degree to which mutations in a heterozygote behave like wildtype) was not constant between environments. Heterozygotes grew stochastically under nystatin stress, and resequencing uncovered rapid and pervasive loss of heterozygosity. Combined, this work demonstrates that both ploidy and the environment can have a large influence on the effect of beneficial mutations and illustrates the often-dynamic nature of evolution.

Preface

A version of Chapter 2 has been published as "Gerstein, A. C., and S. P. Otto, 2011 Cryptic fitness advantage: diploids invade haploid populations despite lacking any apparent advantage as measured by standard fitness assays. *PLoS One* 6: e26599". I conceived of this project in collaboration with S.P. Otto. I performed all laboratory work, performed all analyses and wrote the original manuscript. S.P. Otto provided advice on analyses and contributed revisions to the manuscript.

A version of Chapter 3 has been published as "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". S.P. Otto conceived the project. L.A. Cleathero and M.A. Mandegar performed the experimental evolution experiments under my supervision. S.P. Otto developed the theory presented in Appendix B.3. I conducted the flow cytometry, performed all analyses and wrote the original manuscript. S.P. Otto provided advice on analyses and contributed revisions to the manuscript.

A version of Chapter 4 is currently in press at *Genetics* as "Gerstein, A. C., D.S. Lo, and S. P. Otto, Parallel genetic changes and non-parallel gene-environment interactions characterize the evolution of drug resistance in yeast.". I conceived the project. I performed laboratory work in conjunction with D.S. Lo. S.P. Otto developed the likelihood framework used to analyze the tolerance assays and wrote the scripts to analyze the Illumina sequence data. I wrote the original manuscript. S.P. Otto provided advice on analyses and contributed revisions to the manuscript.

Table of Contents

List of Tables

List of Figures

Acknowledgements

They say that all scientists stand on the shoulders of giants. To my supervisor, Sally, I sincerely thank you for letting me stand on your shoulders all these years. I feel like I won the lottery the day you accepted me into your lab. Everything I know about how to conduct science properly, I learnt from you (I also learnt the difference between then and than). Thank you for your patience, constant support, and unfailing positivity. You have won more awards than cruise ships on the ocean during my tenure in your lab, and I can genuinely say you deserve all of them and more. I can not imagine a better graduate school supervisor.

I would also like to sincerely thank my advisory committee. Vivien Measday opened her door and her lab to me, and I could not have faked my way through yeast genetics without her assistance; thank you for agreeing to be on my committee, even though it was outside your own field of interest, this thesis could not have been completed without you and the members of your lab (in particular, Mike Anderson, Krystina Ho and Nina Piggott). Patricia (Trish) Schulte contributed invaluable scientific knowledge, opinions, and encouragement. I am also greatly indebted to Mike Whitlock, the token male at my committee meetings. Thank you for being the first person in my academic career to make me feel like maybe I could do this thing, and thank you for continuing to be a major source of support. Thank you also for your dedication to making academia a better place for graduate students.

My labmates (and Otto Lab hangers-on) have inspired me scientifically, and helped me to learn some theory without having to do it myself. Thank you Aneil Agrawal, Dilara Ally, Rowan Barrett, Florence Débarre, Rich FitzJohn, Kim Gilbert, Jessica Hill, Kay Hodgins, Liz Kleynhans, Mickael LeGac, Crispin Jordan, Leithen M'Gonigle, Karen Manguson-Ford, Itay Mayrose, Kate Ostevik, Alana Schick, Michael Scott, and Jabus Tyerman. Rich Fitzjohn has improved the R-literacy of the department to an unquantifiable degree, and I am particularly indebted to both Rich and Alistair Blachford for many R, LaTeX and miscellaneous computer-related queries that are too numerous to list. I would also like to thank the tremendous undergraduate students that have worked in the lab with me: Mohammed Mandegar, Lesley Clethero, Laura Glaubrach, William Tang, Gurpreet Khadir, Lisa Ander-

son, Emma Hodgson,Yukon Chen, Sherry Li, William Li, Anna Van Tol, and Tracey Hinder. This thesis truly could not have been completed without their efforts. Dara Lo deserves particular credit. She returned for a second stint in our lab (and then a third and fourth) during a period of my degree when I desperately needed another brain around to help me troubleshoot. Though she was 'just' an undergrad, she more than capably helped me break through the wall. Thanks also to Anastasia Kuzmin for being my sequencing expert (both Illumia and Sanger), and enabling this thesis to be completed in a timely fashion. Thank you also to Jasmine Ono, for finally taking over 'my' lab; I wish you'd started three years ago, but it's been a pleasure to spend my last six months in the lab with you.

The decision to return to UBC Zoology to conduct PhD research was not made easily, yet it was a decision I have not regretted. Some people treat graduate degrees like a job, and spend the majority of their "free" time with friends from outside of graduate school. I was not one of those people. I don't know that I could have made it without the friendship of the people that I've spent my time with inside and out of Biodiv/Biosci. Although a work/life balance is sometimes hard to attain in graduate school, it certainly helps when work and life coincide at department-sanctioned events like Friday Beers, the Huts skit, and miscellaneous ZGSA fun times. I hope these traditions continue long into the future. Thank you to my officemates, Jon Mee and Jean-Sebastién Moore, who stuck by me when nothing was working and I was, I'm sure, the grumpiest person in the building. I'd also like to thank my BFFs, Anne Dalziel, who has been my SG and go-to person for advice for almost my entirety of graduate school, and Milica Mandic, who has been a true ally in life, academia, and on our snowboards. I would also like to thank a number of people who have helped me scientifically, supported me when I needed someone to talk to, and helped make Vancouver a wonderful place to attend graduate school: Mike Barker, Iain Caldwell, Josh Chang, Adam Chateauvert, Gina Conte, Joey Courschene, Will Cornwall, Carla Crossman, Simone Des Roche, Rod Docking, Joel Heath, Christie Hurrell, Nolan Kane, Heather Kharouba, Julie Lee-Yaw, Heather Major, Brook Moyers, Antoine Paccard, Seth Rudman, Kieran Samuk, Graham Scott, Jenny Selgrath, Matt Siegle, Laura Southcott, Ben Speers-Roesch, Dave Toews, Kathryn Turner, Sam Yeaman, John Yensen. Bon Jovi said it best - *"We've got each other, and that's a lot . . . take my hand, we'll make it I swear"*. Woot!

A number of other people and organizations contributed to the success of this thesis. I am truly grateful to the National Science and Engineering Research Council of Canada, the Killam Trusts, and The University of British Columbia for funding me throughout graduate school. I hope that no matter what the future holds, Canada will continue to make independent funding available to students. I would also like to thank Justin and Andy at The UBC Flow Cytometry Facility, for running an excellent facility that enabled much of the research in this thesis. David Pellman welcomed me into his lab at Harvard for two months early on in my degree, and combined with Anna Selmecki and Satoshi Yoshida, taught me a lot

about what it is like to work in a 'real' yeast lab. Barry Williams also invited me to spend three weeks at MSU at the midpoint of my degree; those three weeks did more for my confidence and positive future outlook than anything else could have. I would also like to sincerely thank Mr. Ffrench, my high school biology teacher, for instilling in me the desire to ask and answer questions about biology rather than memorize facts, and who taught me to see the forest instead of the trees. Finally, I would like to thank The Peak radio station for coming into existence when I was spending more than half my waking hours in the lab, and providing an excellent soundtrack that improved my quality of life.

My Masters thesis was dedicated to my parents, Paul and Sherri, and this one could have been as well. I do not think I could have made it this far without you. Thank you for everything you have ever done for me, I could not imagine more supportive parents.

This thesis is dedicated to my brother, Michael,
who has unknowingly been my biggest source of support during this endeavour.
And, to Les: even when I didn't know where to find the tunnel,
I knew you were the light at the end of it.

Chapter 1

Introduction

Benjamin Franklin is famously quoted as saying *"In this world nothing can be said to be certain, except death and taxes"*. True though this statement may be, the process of evolution in biological systems is equally certain. Evolution happens continuously in all populations under all environmental conditions. The biological world is not a static place, and neither are the species nor populations that live within it. Understanding the process of evolution is thus tantamount to understanding life itself.

One of the most remarkable feats of the evolutionary process is that the same mechanism that yields minute differences among populations (e.g., variation in bristle number between *Drosophila melanogaster* individuals, Lai *et al.* 1994) also enables macroevolutionary changes that yield entirely new forms (e.g., plant domestication yielding maize from teosinte, Wang *et al.* 1999). No matter the scope, evolution acts through selection of genetic variants within populations. Newly favourable variants may arise as mutations *de novo* within populations, from previously neutral or deleterious standing genetic variation, or through the migration of alleles from other populations or species, yet at its core, the process of natural selection is entirely limited by the appearance of heritable variation. Although natural selection acts on phenotypes within populations, it is allele frequencies at the genetic level that change over evolutionary time.

Molecular techniques, developed in the 1960s, allowed evolutionary biologists to study variation at the genetic level for the first time. The results that emerged from these studies lead to surprising discoveries about the tremendous degree of variation that exists at the amino acid and protein levels (Nei, 1987). Although much of this variation may be neutral (Kimura, 1968), neutrality alone cannot account for the entirety of within- and between-species variation (McDonald and Kreitman, 1991). A suite of fundamentally important questions remains largely unanswered concerning this variation: does adaptation primarily utilize standing genetic variation or *de novo* mutations (Hermisson and Pennings, 2005;

Barrett and Schulter, 2007)? How many different places in the genome can yield mutations that enable similar phenotypes? Are the pleiotropic and epistatic effects of different mutations that yield the same phenotype in a given environment also the same? How does the environment influence the maintenance of variation?

The genome acts as a palette on which evolution operates, as well as a constraint that limits the myriad pathways that evolution is able to traverse. To comprehensively understand the evolutionary process we must thus understand the genetic basis of adaptive mutations (Orr, 2005). The work described within this thesis combines recent technological advances with an extremely tractable study system (experimental evolution in *Saccharomyces cerevisiae*, a single celled-microorganism) to study the genotypic basis, phenotypic diversity, and fitness effects of beneficial mutations.

1.1 The Use of Microorganisms to Study Adaptive Evolution at the Genetic Level

In 1991 Richard Lenski and colleagues published a landmark paper that described 12 populations of *Escherichia coli* adapted for 2000 generations to minimal-glucose medium (Lenski *et al.*, 1991). Although this was not the first study to describe the use of microorganisms to capture evolutionary processes (e.g., Luria and Delbrück 1943; Dykhuizen and Hartl 1983; Paquin and Adams 1983), it is largely credited with spawning the field of experimental evolution using microbes. In the twenty years that have followed, the 'Lenski' lines have now evolved for over 50 000 cell generations, and these 12 populations have adapted in ways that could not have been predicted at the outset (e.g., obtaining the ability to metabolize citrate as a carbon source, a capacity never before seen in *E. coli*, Blount *et al.* 2008).

A multitude of traits make *E. coli* an attractive species with which to study evolutionary biology. Populations undergo rapidly doublings to grow at extremely large population sizes (upwards of 10^8 individuals/mL), and are easy to grow in high replication (the original experimental lines are evolved in small Erlenmeyer flasks, yet cultures can also be easily grown in test tubes or multi-well plates). Experimental evolution studies are typically initiated by starting many replicate populations of the same (ancestral) genotype, thus this approach has been a fruitful way to explore the range of paths that evolution can take, given the same starting material. Microbial cultures are also amenable to the maintenance of a fossil record, as stocks can be frozen down and maintained at -80°C indefinitely. The ability to regenerate populations from frozen stock is perhaps their most unique attribute among biological systems. This capability has enabled researchers to conduct evolution experiments forward in time, while retaining the capacity to look backwards and trace evo-

lutionary transitions of the specific adaptive alleles that are found to predominate in the population at the end of an experiment. This is in stark contrast to the majority of evolutionary research that is forced to study contemporary populations to infer past mechanisms by which evolution might have happened.

These traits are not unique to *E. coli*, and a diversity of microbial species that span the kingdoms have been similarly utilized by evolutionary biologists. Although these experiments are generally conducted in simple laboratory environments, they maintain a relevance to the real world because researchers do not themselves select the eventual outcome, thus, these populations are subjected to natural (rather than artificial) selection (Dykhuizen, 1990). Experimental evolution has thus been used with tremendous success to study some of the particulars of adaptive mutations that were previously a black box, including questions about the rate at which they appear (e.g., Joseph and Hall 2004; Dickinson 2008; Lynch *et al.* 2008), the distribution of their effect sizes (e.g, Barrett *et al.* 2006; Kassen and Bataillon 2006; Rokyta *et al.* 2008; MacLean and Buckling 2009), and their rate of fixation (including the effects of clonal interference and the maintenance of genetic variation, Desai *et al.* 2007; Kao and Sherlock 2008; Pepin and Wichman 2008).

A final advantage that should not be overlooked in this discussion is the incredibly small genome size of most microorganisms (Gregory, 2005). In part due to its 12.1Mb genome, the budding yeast *Saccharomyces cerevisiae* was the first eukaryote to have a completed genome sequence (Goffeau *et al.*, 1996) (as a point of reference, the human genome is 3.2Gb, while even the puffer fish genome, the smallest known vertebrate genome, is 390Mb). In recent years, with the advent of next generation sequencing technology, it has become feasible to sequence the entire genome of adapted lines in microbes, something that remains considerably out of reach (with respect to both bioinformatics and cost) in species with larger genomes. Whereas previous microbial evolution experiments have shed light on the pattern of evolution, the ability to sequence an entire genome throws open the ability to determine the genetic basis of the mutations that have enabled adaptation. Perhaps unsurprisingly, work done in *E. coli* has again been at the forefront of this research (Barrick *et al.*, 2009; Conrad *et al.*, 2009). The work described in Chapters 4-6 of this thesis would not have been possible without access to this sequencing technology.

All experiments in this thesis were conducted on the budding yeast, *S. cerevisiae*. The decision to use *S. cerevisiae* was twofold. The genomic resources available for *S. cerevisiae* (curated at the Saccharomyces Genome Database, Cherry *et al.* 2011) are without precedent for a eukaryote. I have greatly benefited from the tremendous work that has gone into characterizing the biological and biochemical roles of the $\sim$5800 identified open reading frames ($\sim$ 85% of yeast genes currently have at least one identified biological role, Botstein and Fink 2011), as well as the community of collaboration and knowledge sharing. As I was keenly interested in uncovering the genetic basis of adaptation, a well annotated

genome was a requirement. Beyond the accumulated wealth of knowledge, I also utilized the yeast deletion collections, available sets of lines that each contain a single open reading frame (ORF) deletion for non-essential genes in haploid (Winzeler, 1999) and homozygous diploid (Giaver *et al.*, 2002) backgrounds, and a set of all genes in a heterozygous background (Deutschbauer, 2005). The second reason to utilize *S. cerevisiae* was question driven. As detailed below, many of the major questions explored in this thesis are concerned with the effect of ploidy on adaptation. Working with *S. cerevisiae* allowed me to manipulate the ploidy background of strains and mutations in a relatively easy manner that is not possible in the prokaryotic species that have been commonly used for experimental evolution studies.

1.2 Measuring Fitness in Experimental Evolution Experiments

Evolutionary biologists are tasked with explaining how and why a particular genotype is advantageous in a given environment. This is not easy to determine in any species, as fitness frequently depends on multiple traits expressed throughout the life cycle; the particular importance of different traits may themselves depend on many factors including the abiotic and biotic environments, and the organisms own internal environment. One of the most difficult aspects of working with microbes is our inability to see them. We are not (yet) able to easily follow the offspring of single cells through time, directly track family history, or measure survival at the individual level. As we can not directly measure the reproductive output of single individuals we rely primarily on fitness proxies to tell us which genotypes/phenotypes/ mutants are most likely to contribute the highest number of offspring to future time points.

Experimental evolution studies have utilized two classes of fitness proxies to gain a sense of absolute fitness. The first set of assays measure fitness of lag phase, exponential growth, and stationary phase, the three primary phases of growth that are present during batch culture growth (the type of growth I employed throughout this thesis, Figure 1.1). Measures of stationary phase represent the number of cells that are present after nutrients have been metabolized and growth has largely ceased. Differences between lines of interest in stationary phase measures represent differences in biomass yield from the same starting nutrients. In practice (for populations grown in liquid culture) this is often measured as the change in optical density through a 24 or 48 hour growth cycle, rather than the number of cells directly. Exponential growth measures represent the rate at which a lineage is able to produce offspring when nutrients are not limiting. Growth rate is typically measured using spectrophotometers that track optical density over time across a single growth cycle. Mathematical models are then used to calculate the rate of change during the exponential growth phase. The lag phase at the beginning of the growth cycle is the length of time

that occurs between transfer of cells into fresh medium and the time at which cells begin rapidly growing. This phase is often not measured, because cell growth in this early phase is below the detection limit of the machines we use to measure exponential growth. Lag phase might well be important, however, if differences exist between lineages in the ability to begin growing once nutrients in their environment become available.

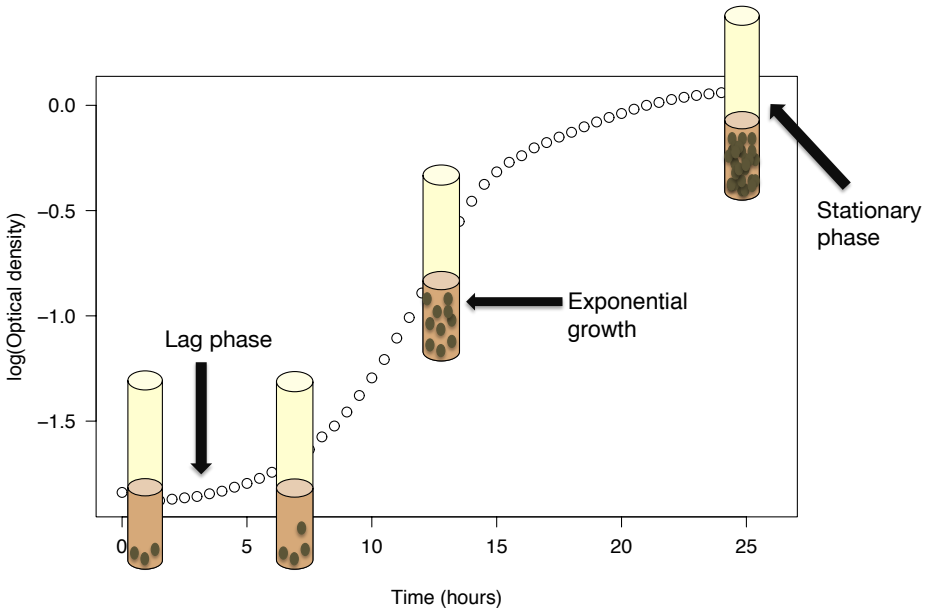


Figure 1.1: The three primary phases of batch culture growth: lag phase, exponential growth and stationary phase. Points represent the log of optical density measurements taken through a typical growth cycle from the time cells are inoculated into fresh medium until nutrients have been utilized. The cartoon of test tubes is used to represent how the number of cells relates to optical density.

Competitive fitness assays that directly compare two (or more) strains against each other are the second class of fitness proxies used to examine the fitness of a line (or lines) of interest. Many evolutionary experiments assay competitive fitness by competing evolved strains directly against an ancestor strain that differs by a phenotypic marker (e.g., auxotrophic or fluorescence markers). Competition experiments are generally carried out in a manner that mimics the way in which strains were evolved. These experiments are often initiated with equal numbers of the ancestral and evolved individuals, though the starting ratio can be altered and may sometimes have an effect on competitive outcome (discussed in detail in the Chapter 7). After some number of generations (generally between 7-100), the ratio of competing strains is determined, and the change in ratio from the beginning to end of the competition experiment indicates the competitive fitness of the strain of interest. A competition experiment can also be used to simultaneously compare a large number of strains if

many markers are available. The yeast deletion set of lines (where each line carries a single gene deletion in the same ancestral background) was constructed such that each gene knockout carries a unique barcode. The entire set of lines can thus be competed against each other, with the change in frequency of each barcode (representing each gene knockout) determined at the start and end of a given experiment to determine the fitness of each line relative to the others.

In the second chapter of this thesis I compared many different fitness assays. Previous work I did during my Masters (Gerstein *et al.*, 2006) found that diploid mutants arose and overtook haploid individuals during 1800 generations of batch culture evolution. In Chapter 2 I conducted a series of fitness assays with the aim of determining what aspect(s) of fitness enabled diploids to outcompete haploids. We looked in-depth at one initially haploid line that showed this ploidy transition; we first isolated many colonies from across the timeseries and determined the ploidy of each. We then focused on two different comparisons – we looked at a small number of colonies isolated across the timeseries to examine which traits changed over the ∼1800 generation experiment, and secondly we compared haploid and diploid colonies isolated from within the same population at 1023 and 1302 generations, the two time points where both haploids and diploids were common. We measured fitness during all three major phases of growth as well as competitive fitness against a common competitor (closely related to the ancestor). We also developed an assay to compete colonies of different ploidy directly against each other, using ploidy as the marker.

The intellectual contribution from this project towards the rest of the thesis was important. As described below, understanding ploidy is a major driver of many of the experiments I conducted within this thesis, and this project initiated a lot of thought on what might be different between haploid and diploid cells. Furthermore, the extensive set of experiments conducted in Chapter 2 demonstrated that in many cases different fitness assays may yield different insights into factors that underly important differences between genotypes of interest. This has also been shown in a mutation accumulation study with *S. cerevisiae* that found mutations that altered one component of fitness generally had little effect on other components (Hall and Joseph, 2010).

For the remainder of the thesis I used multiple fitness assays wherever possible, and spent time thinking about which assays best fit the questions we were trying to answer. In Chapters 3-6 I conducted three distinct types of fitness assays. We assayed competitive fitness against a common competitor using flow cytometry to measure how fitness changed between ancestral and evolved populations. This allowed us to assess the ability of one genotype to overtake another in the same arena within which populations were adapting. I also used growth curve experiments to measure both intrinsic growth rate and biomass production. Lastly, I used dose-response curves, to assess the tolerance breadth of genotypes of interest. This assay is unique among those I utilize, as it does not require us to

pick a single level of stressor to assay, but rather allows us to determine how cells respond to an increasing strength of a given stressor. From these data we calculate the half maximal inhibitory concentration of a stressor (IC_{50}), as well as the slope of growth decline and the asymptote of growth at low levels of stressor. All assays measured fitness during clonal growth. Combined, these assays allow us to compare multiple fitness axis among our genotypes of interest.

1.3 Ploidy

The effect of ploidy on evolution is one of the overarching themes of this thesis. Ploidy has the potential to greatly affect the trajectory of evolution, yet the factors that have contributed to the variation in ploidy seen throughout the tree of life remain poorly understood. Natural isolates of varying ploidy have been identified for the species I focus on, *S. cerevisiae* (Ezov *et al.*, 2006). Isogenic ploidy variants of *S. cerevisiae* are known to differ in a wide variety of aspects including cell size and the surface area to volume ratio (Galitski *et al.*, 1999; Mable, 2001), protein levels (de Godoy *et al.*, 2008), and gene expression (Wu *et al.*, 2010). Furthermore, from an evolutionary standpoint, many of the factors that potentially differ between populations of varying ploidy may directly influence the appearance and spread of beneficial mutations (Zeyl *et al.*, 2003).

Evolution occurs when a novel adaptive variant appears in a population and rises in frequency. The rate at which new mutations appear is a function of both the population size and the mutation rate, as either parameter increases, so too does the appearance of new mutations. Similarly, the fixation of new mutations, once they are in the population, depends on their effect size, as mutations of larger effect fix more rapidly than mutations of smaller effect. Although theoretical studies often assume that mutations in haploid and homozygous diploids have the same effects, very few empirical studies have directly examined this question. If the same mutations have different effect sizes in haploids and diploids, this could directly enable differences in the rate of adaptation between populations of varying ploidy. The effect size of mutations in diploids may also be partially masked when a wild-type allele is also present in the genome. In diploid individuals, mutations generally arise in heterozygous form (i.e., in a single copy), and they will remain so in asexual lineages until loss of heterozygosity occurs. This fundamental difference between mutations that appear unmasked in haploids but in heterozygous form in diploids has the potential to drastically slow the fixation rate of adaptive mutations in diploids, as the beneficial effects of new mutations may be partially or entirely masked by the remaining wild type allele. If mutations are primarily recessive in a given environment (i.e., if the fitness of a heterozygote is more similar to wildtype than a homozygous double mutant), haploids are predicted to evolve much faster than diploids. If semidominant or dominant mutations are available, however,

then diploids (with twice the mutational targets) should adapt faster than haploids when mutations are limiting (Orr and Otto, 1994).

The haploid:diploid rate of adaptation thus depends on the rate that new mutations appear in the population, their effect sizes, and their dominance values. At large population sizes, where mutations are not limiting, haploids are expected to adapt faster than diploids, a prediction that has been empirically verified by comparing haploid and diploid populations adapting in *S. cerevisiae* (Zeyl *et al.*, 2003), and *Aspergillus nidulans* (Schoustra *et al.*, 2006). At smaller population sizes, when mutations were limited, the difference in the rate of adaptation disappeared (Zeyl *et al.*, 2003). The dominance of mutations has also been empirically shown to drive differences in the rate of adaptation. When haploid and diploid populations of yeast were evolved at a concentration of fluconazole where semi-dominant mutations were previously shown to exist (Anderson *et al.*, 2003), diploids were able to acquire mutations and adapt faster than haploids. When the diploid population size was halved, or the haploid population size doubled, this advantage disappeared and mutations were acquired at the same rate (Anderson *et al.*, 2004). At higher concentrations of fluconazole, where dominant or semidominant mutations do not appear to be available, haploids adapted much faster than diploid populations.

Differences in the rate of adaptation can thus arise due to environment- and/or ploidy-specific differences in the mutation rate, mutation effect sizes, or variance in the dominance properties of adaptive mutations. In Chapter 3 I directly compared the rate of adaptation between haploid and diploid populations evolved in seven different environments to gain a sense of how the environment and ploidy combine to influence the rate of adaptation. In Chapter 5 I compared the effect size of twenty mutations in haploid and homozygous diploids, while in Chapter 6 I compared the same mutations in heterozygous and homozygous diploids to gain a sense of their dominance. Without direct empirical measurements, it is difficult to disentangle the confounding effects of dominance and effect size in diploids. If, for example, mutations have lower effects in diploid compared to haploids, it may appear that mutations are more recessive than they truly are if heterozygous fitness effects are compared to haploid fitness effects.

1.4 The Genotypic Basis of Adaptive Mutations

Although evolutionary biologists have long been interested in characterizing adaptive mutational variants at the genotypic level (e.g., Cuevas *et al.* 2002; Harr *et al.* 2002; Cooper *et al.* 2003), the precise mutational events that lead to adaptation remain poorly understood. In particular, the number of different avenues available for adaptation to a given environment remains unknown (though, as with much of evolution, the answer likely depends heavily on the species and environment in question). Repeated evolution, defined as

the independent appearance of similar phenotypic traits in distinct evolutionary lineages (Gompel and Prud'homme, 2009) has been consistently 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* during lung colonization of 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 these observations of parallel phenotypic change. Recent reviews have compiled many examples of convergent phenotypic changes that both are and are not the result of parallel genetic changes (Manceau *et al.*, 2010; Christin *et al.*, 2010; Gompel and Prud'homme, 2009). What remains unknown from these studies, however, is the evolutionary history of parallelism (or lack of parallelism). Without knowledge of this history, it remains difficult to infer whether convergent evolution of phenotypic and genotypic changes is the result of constraint or selection (Losos, 2011). Furthermore, although contemporary species may show any particular pattern (e.g., convergent phenotypes with a different genetic basis), it remains difficult if not impossible to determine whether the evolutionary trajectories were parallel from the start or initially took place in different directions, converging over time.

Experimental evolution studies have (perhaps unsurprisingly) become a fruitful avenue of research to characterize the genetic basis of adaptation (Conrad *et al.*, 2011). Sequencing replicate populations initiated with the same ancestral culture has demonstrated that in many cases the same genes repeatedly acquire beneficial mutations. One of the first studies to look for parallel changes examined four *Saccharomyces cerevisiae* lines evolved to glucose-galactose fluctuating environments across 36 sexual cycles (Segrè *et al.*, 2006). With targeted sequencing they found that all four lines had acquired mutations in *GAL80*; the exact same nucleotide was mutated in two different lines. A similar result was found by Applebee *et al.* (2011), who evolved 50 *E. coli* lineages on glycerol minimal medium for 24 or 40 days and sequenced *glpK*, the only gene that was previously found to evolve in all five previously examined strains (Herring *et al.*, 2006). They identified single non-synonymous *glpK* mutations in 47 out of 50 lines. Five codons were repeatedly mutated, with one residue mutated in 20 of the 50 lineages (the initial amino acid, aspartic acid, was changed to valine in 7 lineages and alanine in 13 lineages). Barrick *et al.* (2009) also found evidence that the same genes are repeatedly targeted throughout independent adaptation to the same environment. Of 14 genes that were re-sequenced in 12 *E. coli* lines independently evolved in a glucose limited environment for 20 000 generations, three genes were mutated in all populations, and nine additional genes carried mutations in multiple lines (between 3 and 8 lines); in almost all cases different mutations were implicated (Barrick *et al.*, 2009; Woods *et al.*, 2006; Pelosi *et al.*, 2006; Cooper *et al.*, 2003). Mutations other than SNPs also repeatedly arise during experimental evolution experiments, for example, Chou *et al.* (2009)

documented a transposition mutation repeatedly arising in 30/32 lines of *Methylobacterium* evolved for 1500 generations on methanol.

The magnitude of genetic parallelism among lines is likely to be heavily environmentally dependent. Gresham *et al.* (2008) found that 8/8 *S. cerevisiae* lineages evolved to a sulfate-limited environment had duplicated the sulfur transporter gene *SUL1*. In contrast, lines evolved to glucose- or phosphate-limited environments showed a much larger diversity of selected mutations within the same timeframe. The difference between environments may lie in the fitness surface; in cases with a specific challenge involving a narrow fitness ridge, the genomic target for beneficial mutations might be small. Alternatively, the difference between environments might lie in the probability of establishment of different mutations; if some mutations have a much larger benefit than others, then we would expect such mutations to arise and outcompete other mutations in large microbial populations (indeed, *SUL1* duplications increased fitness by 50%, Gresham *et al.* 2008). Epistatic relationships between adaptive mutations might also influence observed parallelism. As shown by Anderson *et al.* (2010) and Parreiras *et al.* (2011), antagonistic epistasis could well limit the number of routes available to adaptation by restricting the number of available mutations.

In Chapter 4 I focused on characterizing the genotypic and phenotypic basis of 35 mutation lines I acquired upon exposing initially haploid and isogenic yeast to the fungicide nystatin. The lines were evolved for a very short period of time, such that we expected very few beneficial mutations to be present in each line (which minimizes the possibly confounding effect of epistasis, as mutation lines were isolated at the first indication of growth). This chapter is thus unique among recent studies in fully analyzing the first mutational steps in a eukaryote, using whole genome resequencing. Understanding the first steps of adaptation to a novel environment provides insight into the diversity of mutational avenues that are available to populations when first presented with a stressful environment.

1.5 Summary

The work done for this thesis has contributed to our understanding of how evolution works at the genomic level. Through the characterization of beneficial mutations I have examined the appearance of variation at multiple biological levels. In Chapter 2 I describe a case study experiment, where we have looked for differences in fitness between haploid and diploid genotypes that were isolated from the same population. Although we compare across many different assays (including growth phase and competitive fitness) we did not find any parameters for which diploids showed a clear advantage. This was interesting, as we had observed clear and repeated diploid takeover in previous initial experiments. In Chapter 3 I compared the rate of adaptation between haploid and diploid populations across seven

different environments, using a competitive fitness assay. We found that both ploidy and environment significantly influenced the rate of adaptation, and discussed the factors that might lead to population level variation. Chapters 4-6 are based on a set of mutation lines I acquired that confer tolerance to nystatin, a fungicide. In Chapter 4 we determined the genetic basis of adaptation for all mutation lines. We demonstrated that although the genetic basis of adaptation was narrow, lines that carried mutations in different genes exhibited strikingly different trade-offs under a range of environmental pressures. This work uncovered unpredictable phenotypic variation between individuals that carry mutations in genes that act close together within a regulatory pathway. I compared the effect size of the twenty unique nystatin mutations in haploid and homozygous diploid backgrounds in Chapter 5 and found that mutations in a haploid background had larger effect sizes than the same mutations in homozygous diploids. When these same mutations were placed in a heterozygous diploid background we found that heterozygotes grow much more stochastically than haploids or homozygous diploids (Chapter 6). Resequencing some heterozygous replicates grown under stressful conditions demonstrated rampant and rapid loss of heterozygosity. Combined, this thesis demonstrates the dynamic nature of the evolutionary process, and advances our understanding of how ploidy can influence the effects of beneficial mutations.

Chapter 2

Cryptic fitness advantage: diploids invade haploid populations despite lacking any apparent advantage as measured by standard fitness assays

2.1 Summary

Ploidy varies tremendously within and between species, yet the factors that influence when or why ploidy variants are adaptive remain poorly understood. Our previous work found that diploid individuals repeatedly arose within ten replicate haploid populations of *Saccharomyces cerevisiae*, and in each case we witnessed diploid takeover within ~1800 asexual generations of batch culture evolution in the lab. The character that allowed diploids to rise in frequency within haploid populations remains unknown. Here we present a number of experiments conducted with the goal to determine what this trait (or traits) might have been. Experiments were conducted both by sampling a small number of colonies from the stocks frozen every two weeks (~93 generations) during the original experiment, as well as through sampling a larger number of colonies at the two time points where polymorphism for ploidy was most prevalent. Surprisingly, none of our fitness component measures (lag phase, growth rate, biomass production) indicated an advantage to diploidy. Similarly, competition assays against a common competitor and direct competition between haploid and diploid colonies isolated from the same time point failed to indicate a diploid advantage. Furthermore, we uncovered a tremendous amount of trait variation among colonies of the same ploidy level. Only late-appearing diploids showed a competitive advantage over haploids, indicating that the fitness advantage that allowed eventual takeover was not

diploidy per se but an attribute of a subset of diploid lineages. Nevertheless, the initial rise in diploids to intermediate frequency cannot be explained by any of the fitness measures used; we suggest that the resolution to this mystery is negative frequency-dependent selection, which is ignored in the standard fitness measures used.

2.2 Introduction

The study of adaptive evolution is in many ways the study of fitness. Having identified an interesting pattern in nature, we examine and compare fitness differences within contemporary populations to infer how evolution might have happened. This method has notoriously been criticized by one of the most widely cited papers in the field (Gould and Lewontin, 1979) because we traditionally lack the ability to perform direct experiments on the individuals that were actually present at the time when evolution occurred to determine which mutation provided an advantage (and why). Recently, experimental evolution with microbes has provided an approach whereby the entire process of evolutionary change can be studied and used to test adaptive processes directly, without inference about the populations and individuals involved.

Our previous work reported a surprising result that arose during an $\sim$ 1800 generation batch culture evolution experiment. We found that diploid individuals arose within haploid populations of *Saccharomyces cerevisiae* and eventually swept independently in ten lines, even though the lines were asexual (5/5 lines evolved in YPD & 5/5 lines evolved in YPD+salt, Gerstein *et al.* 2006). We proposed that historical contingency may be acting; as *S. cerevisiae* is historically diploid, selection may have acted on rare diploid individuals that arise naturally at low frequency to regain this historical state. However, the true character on which selection was acting to allow diploids to take over remains unknown. Here, we present experiments conducted with the goal of determining what fitness component allowed diploids to repeatedly invade haploid populations.

The question of why one ploidy level is able to outperform another over evolutionary time is of broad interest, as tremendous variation in ploidy is seen throughout the tree of life, even among closely related species (Parfrey *et al.*, 2008). All sexual species undergo a ploidy cycle over a generation, and some species maintain prominent haploid and diploid free-living stages (i.e., alternation of generations), while in other species ploidy differs between sexes (i.e., haplodiploidy, e.g., in Hymenoptera). Though ploidy variation is pervasive, we generally have a poor understanding of when or why ploidy variants are adaptive. As one example, a recent study that examined the link between ploidy and plant species worldwide found endangered plants were disproportionately diploid, while invasive species were more likely to be polyploid (Pandit and Pocock, 2011), yet the traits that underlie these correlations remain unknown.

The species we focus on, *Saccharomcyes cerevisiae*, is itself known to display multiple ploidy levels in natural isolates (Ezov *et al.*, 2006). Ploidy variants of *S. cerevisiae* are known to differ in a wide variety of aspects, even when isogenic. Cell size increases as ploidy increases (Galitski *et al.*, 1999; Mable, 2001); as *S. cerevisiae* cells are prolate spheroids, increases in volume decrease the surface area to volume ratio, and thus diploids, which are typically larger than haploids, have a significantly lower surface area to volume ratio. Gene expression and protein levels also differ between isogenic individuals with different ploidy; de Godoy *et al.* (2008) found that 196 proteins changed more than 50% in abundance between haploids and diploids, while Wu *et al.* (2010) recently observed that 65 genes differ in expression between haploid and tetraploid *S. cerevisiae* isogenic individuals. Interestingly, it may be cell size rather than ploidy that influences gene expression, as Wu *et al.* also showed that genes with expression differences between haploids and tetraploids also differed in expression (in the same direction) when comparing wild type haploids with *cln3*Δ haploids that are 185% the volume of wildtype haploids.

Previous researchers have found mixed results when comparing haploid and diploid fitness under conditions similar to ours (isogenic haploids and diploids grown in rich medium at 30°C). Adams and Hansche (1974) and Temina *et al.* (1979) found no difference in growth rate between haploids and diploids under these conditions, while Mable (2001) found haploids grew significantly faster than diploids (though this was not significant after correcting for multiple comparisons). Similarly, although Glazunov *et al.* (1989) found diploids outcompeted haploids, Mable (2001) found haploids and diploids competed equally well against a common haploid or diploid competitor in YPD at 30°C. Overall, previous work in our lab and others has failed to identify any distinct fitness advantage of diploids over haploids under the conditions of our long-term experiment.

Here we set out specifically to determine why diploid individuals, when they arise by chance within haploid populations, were able to out perform haploids under our experimental conditions. To identify the character that might have allowed diploids to overtake haploids, we conducted a set of fitness assays on haploid and diploid individuals isolated at regular intervals throughout the time series of the original experiment (1767 generations). We can thus assess how different fitness components changed throughout the duration of the experiment. We also conducted competitive fitness assays by directly competing haploid and diploid individuals isolated from the same time point. We focus our attention on two time points in particular, where the diploids have recently risen to appreciable frequency (appreciable enough to be sampled), suggesting a recent selective advantage, and the last time point where haploid colonies were sampled.

We first assayed cell size and shape of haploid and diploid colonies isolated throughout the time series to gain a sense of the magnitude of phenotypic change. We then compared fitness between haploid and diploid genotypes in a variety of ways. It was important to

assay many possible aspects of total fitness, as a previous study that acquired mutations through mutation accumulation in *S. cerevisiae* for 1012 generations found that mutations that altered one component of fitness generally had little effect on other components (Hall and Joseph, 2010). We thus assayed colonies for three fitness components that correlate to the three main phases of growth during batch culture, i.e., lag time upon entering fresh medium, growth rate during logarithmic growth, and biomass production (yield) after 24 hours of growth (transfer into fresh medium was done every 24 hours in our original experiment). We then conducted two types of competition assays: in the first we compete all individuals of interest against a common competitor (closely related to the ancestor), in the second we developed a novel assay that allowed us to directly compete haploid and diploid genotypes isolated from the same time point against each other. We found, surprisingly, that none of these assays indicate a clear diploid advantage that could explain how diploid genotypes were able to rise in frequency within the initially haploid populations. One possibility is that the initial rise in the frequency of diploidy was due to an aspect of fitness not measured by any of these metrics, and we suggest that frequency-dependence may be involved. We also observed significant variation among colonies of the same ploidy level isolated at the same time point. The eventual fixation of diploidy involved a strain that did exhibit a competitive advantage, suggesting that only a subset of diploids could rise to fixation. We conclude that although the end evolutionary result may be deterministic (i.e., that diploids repeatedly take over the population) the route to takeover appears to be largely stochastic, depending on the exact genotypes that arise.

2.3 Materials and Methods

2.3.1 *Isolating ploidy variants*

We previously reported the convergence of 10 replicate haploid lines towards diploidy during ∼1800 generations of batch culture evolution (1767 generations total) (Gerstein *et al.*, 2006). The ancestral strain haplotype is *MATa-a1 ste6△8-694 ura3△0 leu2△0 his4△0 trp1△0 can1△0*. The mutation at the *MAT* locus and *STE6* partial deletion should ensure complete asexuality; previous work found no evidence of revertants at the *MAT* locus or evidence of sexual reproduction (Gerstein *et al.*, 2006). This past work reported a snapshot of genome size change, by assaying the ploidy level of only a single colony from each of the 10 lines every 93 generations up to generation 744, and a single colony from each line at generation 1767. These colonies were obtained from stocks frozen every two weeks in the original experiment (corresponding to $\text{Log}_2 101^{*}14 \sim 93$ generations with daily 1:101 dilutions in batch culture). Here we focus on only one line ("Line A") that had been grown in YPD (an unstressful medium) and showed complete diploid takeover by generation 1488 (Figure A.1),

but analyze multiple colonies from multiple time points to gain a more complete sense of the relative number and fitness of haploid and diploid individuals.

We first isolated ploidy variants from throughout the time series. Freezer stock acquired during the initial evolution experiment (Gerstein *et al.*, 2006) was streaked onto YPD plates and allowed to grow for 72 hours. 24 colonies from each time point were haphazardly picked, inoculated into 10mL YPD and allowed to grow overnight. Flow cytometry on a FACSCalibur was performed as previously described (Gerstein *et al.*, 2006) to assay the ploidy of each colony. Culture from these isolated colonies of known ploidy were then frozen in 15% glycerol for use in all later experiments. We found extensive polymorphism for ploidy from generation 744 to generation 1302 (Results, Figure 2.1), which allowed us to undertake the experiments described below.

We designed a number of experiments to determine whether the ploidy of a sampled colony directly influenced different components of fitness. We first assayed a small number of colonies (two to four) from approximately every 93 generations over the entire 1767 generation time series. These data allowed us to ask whether the fitness measures changed over evolutionary time during the evolution experiment and whether haploid and diploid colonies responded differently. We then took a more in depth look at colonies from generation 1023 and generation 1302. These time points were chosen specifically, as these are the first and last timepoints where polymorphism for genome size was prevalent, and we believed that they would shed the most light on the relative fitness advantage that allowed diploids to invade. We randomly picked five haploid and five diploid colonies (of the 24 initial colonies assayed) from each of these time points and used these same twenty colonies (2 timepoints $\times$ (5 haploid colonies + 5 diploid colonies)) for all subsequent assays. We should note that we do not know whether multiple colonies isolated at the same time point are different genotypes, how many times diploid colonies independently arose, or whether colonies isolated at later generations are the direct descendants of colonies isolated earlier.

2.3.2 *Cell size and shape*

We first conducted an imaging experiment to measure the cell size and shape of haploid and diploid colonies isolated throughout the time series. As these parameters are known to differ between cells of different ploidy, they may directly contribute to fitness differences, as well as indicate the magnitude to which evolution acted within 1767 generations. We assayed colonies from across the entire time series and from twenty colonies isolated at generations 1023 and 1302. The imaging experiment was initiated by streaking colonies onto plates from freezer stock kept at -80°C and allowed to grow for 72 hours. One colony from each line of interest was then randomly picked, inoculated into 10mL of YPD and grown shaking at 30°C for 24 hours. One slide was prepared from each culture using stan-

dard practices. A Zeiss Axioplan microscope with a digital camera attached was used to take at least three digital pictures of each slide (see Figure A.2 for representative haploid and diploid images). Fifteen individual cells were randomly chosen from across the pictures (any cell touching another cell or in the process of budding was excluded). Using the software ImageJ (available at http://rsb.info.nih.gov/ij; developed by Wayne Rasband, National Institutes of Health, Bethesda, MD), photos were enhanced and ellipses were manually drawn around the perimeter of each chosen cell to obtain a length measurement (major axis, L) and width measurement (minor axis, W). We calculated two cell size parameters using the appropriate equations for prolate spheroids, volume (V) and surface area (SA):

$$V = \frac{4}{3}\pi\frac{L}{2}\left(\frac{W}{2}\right)^2 \tag{2.1}$$

$$SA = 2\pi\frac{W}{2}\left(\frac{W}{2} + L\frac{arcsin(e)}{e}\right) \tag{2.2}$$

The equation for surface area depends on the measure of eccentricity (e), which we also used as a descriptor of cell shape:

$$e = \sqrt{1 - \frac{(W/2)^2}{(L/2)^2}} \tag{2.3}$$

Lastly, we calculated the surface area to volume ratio (SA:V), which also describes a component of cell shape.

2.3.3 *Fitness components*

We first measured various fitness components to test whether there were consistent and significant differences over time and between haploid and diploid colonies. We picked three fitness assays that largely reflect the different phases of *S. cerevisiae* growth in YPD during the 24 hours between transfers in our primary experiment. A brief lag phase occurs after transfer into fresh medium, before cells begin growing, followed by a phase of exponential growth during which *S. cerevisiae* rapidly grow and reproduce by fermenting glucose. A diauxic shift between glucose fermentation and ethanol respiration typically occurs around 20 hours for wildtype cells grown in YPD (Stahl *et al.*, 2004). During this postdiauxic phase *S. cerevisiae* grows much slower by respiring the ethanol that is a byproduct of glucose fermentation. As transfers are done every 24 hours we expect growth in this last phase to be under weaker selection (but such growth could contribute to biomass production measured at 24 hours).

Lag phase. To determine the growth lag, we measured the rate at which glucose was consumed by HPLC. We could not use automated optical density measures (OD, see below), because growth during lag phase occurred below the detection limit of our bioscreen machines. For HPLC, we measured two independently cultured replicates of the ancestral ("Geno") haploid, evolved diploid ("Gen1767"), and three haploid and three diploid colonies isolated from 1302 generations (these colonies were a subset of the haploid and diploid colonies isolated from this time point used in all other experiments). For each, a small amount of previously frozen culture was inoculated into 10mL YPD and grown for 48 hours, shaking at 30°C. Five replicate test tubes were then inoculated with 100μL from each culture. At precisely 2, 4, 6, 8 and 24 hours, one replicate tube for each colony was removed from the incubator. Tubes were thoroughly vortexted and 2mL aliquots were pelleted. 1mL of liquid from each tube was filtered with a 25mm filter into a sterile culture vial. Vials were kept at 4°C until the end of the experiment (24 hours). Samples were then run on an Agilent 1100 Series LC/MSD with a Nuleogel Ion 300 OA column at 71°C. The solvent was 4.25 mM H_2SO_4, run isocratically at 0.55 ml/min. Glucose was detected and quantified with a refractive index detector running at 40°C, where the reduction of glucose levels during the earliest time points reflects growth during lag phase.

Growth rate. Naively, one might expect that diploid mutants overtake haploids because diploids grow faster during log growth, which we tested in two sets of experiments, one which examined the small number of colonies isolated throughout the time series and a second that examined 20 colonies isolated from 1023 and 1302 generations. Growth rates were determined using the Bioscreen C Microbiological Workstation (Thermo Labsystems), which measures optical density (OD) in 100 well honeycomb plates, with constant shaking and temperature. Previous work had found growth rate can be variable across bioscreen runs (likely due to small differences in medium, Chapter 7). As these two sets of experiments were not conducted at the same time we compare results only within a single bioscreen experiment. Plates were streaked from frozen stock and allowed to grow for 72 hours. An inoculation containing multiple colonies was allowed to grow overnight in 10mL YPD. 100 μL was transferred into 10mL of fresh YPD, mixed well, and seven 150 μL aliquots from each test tube were placed into different bioscreen wells.

Order of wells was fully randomized. Plates were kept in the Bioscreen C at 30°C, with OD readings taken every 15 minutes for 48 hours. The maximal growth rate was determined for each well as the spline with the highest slope, from a loess fit through log-transformed optical density data (analysis program written by Richard Fitzjohn in the R programming language, R Development Core Team 2011, as previously described, Gerstein and Otto 2009). We interpret this slope as the maximum growth rate in each bioscreen well (which we refer to as "growth rate" throughout).

Biomass production and number of cells at 24 hours. The ability to convert nutrients in the medium into cellular material may also differ over time or between haploids and diploids. We interpret the optical density (OD) at 24 hours as a measure of total biomass production between transfers. For each bioscreen well, we calculated the optical density at 24 hours minus the optical density at the start of the experiment. As haploid and diploid cells (and cells of different genotypes) may differ in cell size, differences in biomass production do not necessarily correlate to differences in absolute number of cells, and we avoid interpreting them as such. To obtain a measure of the number of cells present at 24 hours we conducted hemocytometer counts of ancestral (Geno) and evolved (Gen1767) culture, as well as the five haploid and five diploid colonies isolated after 1023 and 1302 generations of evolution. We note that both growth rate and biomass production were measured in a different environment than the original experiment (100 well honeycomb plates versus large test tubes), and it is possible that a different result could have been obtained if we examined these parameters in the evolutionary environment. Experiments in our lab (unpublished results, A.C.G.) have found very little difference between the parameters measured in these different environments, though we did not test the lines of specific interest for this project.

2.3.4 *Competition against a common competitor.*

A competition experiment was undertaken to gain a comprehensive measure of the ability of each line to compete for resources. A common competitor was created as previously described (Gerstein *et al.*, 2011). Briefly, we inserted a 3320-bp region of the pJHK043 plasmid (generously provided by John Koschwanez, FAS Center for Systems Biology, Harvard University) containing YFP under control of the *ACT1* promoter linked to a histidine marker into the *HIS* locus of BY4741 (*MATa his3*$\triangle$*1 leu2*$\triangle$*0 met15*$\triangle$*0 ura3*$\triangle$*0*), obtained from Open Biosystems (Thermo Fisher Scientific). Competitive fitness was measured by tracking the ratio over time of fluorescing (competitor) cells to the non-fluorescing cells of interest. We measured the competitive fitness of ancestral and evolved culture, and twenty haploid and diploid colonies isolated from generations 1023 and 1302.

For each line of interest culture was streaked from frozen onto YPD plates and grown for 72 hours at 30°C, at which point colonies were inoculated into 10mL YPD and grown for 48 hours. To start the competition experiment, 100μL of the competitor and 100μL of the strain of interest were inoculated into 10mL of YPD. Four replicates were initiated for each line of interest. We performed transfers that exactly mimicked the initial evolution experiment (100 μL transferred from each tube after 24 hours of growth into 10mL fresh medium in large test tubes) for three days. Each day (including the initial day of the experiment), exactly two hours after transfer, 1mL of culture from each tube was aliquoted into an eppendorf, pelleted, and resuspended in 1mL of sodium citrate. 150μL from each

eppendorf was aliquoted into one well of a 96 well plate and immediately run on an LSRII flow cytometer with the High Throughput Sampler attachment. 10000 cells were measured from each well.

Data was analyzed in FlowJo version 8.7 (Tree Star, Inc.). Small debris was excluded with an initial gate then gates were drawn around the two clusters of non-fluorescing and fluorescing cells, by examining plots of FITC-A and AmCyan-A. Clusters were always easily distinguished. The absolute number of cells in each gate for each day of the experiment was determined. The competitive fitness (m) was determined for each line using the formula for evolutionary change:

$$NonFluor = \frac{p_0 e^{mT}}{1 - p_0 + p_0 e^{mT}} \tag{2.4}$$

where *NonFluor* is the fraction of non-fluroescing cells, p_0 is the initial fraction of non-fluorescing cells at the start of the experiment, T is the generation number (measurements were done on days 0, 1, 2, and 3 which corresponds to 0, 6.7, 13.2, and 20.0 generations) and m is the Malthusian parameter of the experimental strain minus that for the YFP-marked competitor (relative growth rate). We use the nls function in the R programming language (R Development Core Team, 2011) to determine the best fitting p_0 and m for each competition assay.

2.3.5 *Direct competition between haploids and diploids*

We also directly competed contemporaneous haploid and diploid colonies isolated from the original experiment to test whether diploid colonies isolated at 1302 or 1488 generations (the latter is the first generation where only diploid colonies were sampled) were able to outcompete a population of the haploids colonies isolated at 1302 generations. All 24 colonies originally isolated at 1302 and 1488 generations were streaked again to single colony from freezer stock maintained at -80°C. A single colony from each was inoculated into 10mL YPD and allowed to grow overnight. 20μL from each of the six haploid colonies isolated from 1302 generations were combined in a single 10mL tube and grown together for a second night, forming the haploid competitor. Two diploid populations were similarly created from diploid colonies isolated at 1302 and 1488 generations by inoculating 10μL of 24 hour cell culture from each of the appropriate diploid colonies into 10mL YPD. The next day, two replicate competitions for each single diploid line and two replicate competitions for each diploid population were initiated by combining 50μL of the haploid competitor and 50μL of diploid culture into 10mL YPD. 100μL was transferred into fresh medium every 24 hours for the next 14 days, exactly mimicking the original evolution experiment. The

initial (day 0) tubes were kept at 4°C for the duration of the experiment. Day 14 tubes were also kept at 4°C after the completion of the experiment until we were able to assay them as described below.

An assay based on standard yeast lab protocols was developed to differentiate between haploid and diploid cells using hydroxyurea, a drug that arrests yeast cells during DNA synthesis (Slater, 1973). Our usual flow cytometry protocol measures cells at all stages of the life cycle, thus haploid cells in G2 have the same DNA content as diploid cells in G1. Arresting cells in S phase, however, allows us to discriminate between haploid and diploid colonies. Seven days after the last transfer, 100μL from all day 0 and day 14 tubes were transferred into fresh medium and grown overnight. The next day, 100μL from each tube was again transferred into fresh medium and allowed to grow for four hours. 1mL from each tube was then added to 200μL of 1M hydroxyurea and transfered into eppendorfs. Eppendorfs were laid flat in a shaking 30°C incubator for 3 hours. We then used the flow cytometry protocol previously described to assess ploidy using a FACSCalibur (Gerstein *et al.*, 2006). Culture was pelleted, resuspended in cold 70% ethanol, and kept at room temperature overnight. The next day culture was pelleted and resuspended twice in 1mL sodium citrate, 25μL RNAse A was added and tubes were incubated at 37°C overnight. Tubes were again pelleted and resuspended twice in 1mL sodium citrate. 30μL sytox green was added and tubes were left at room temperature in the dark overnight. The next morning all samples were run on a FACSCalibur. Although this method does not perfectly assay ploidy level (some cells escape arrest, see Figure A.3) we found the fraction of un-arrested cells to be fairly consistent. We focus on the change in the frequency of diploid cells initially and after 14 days of competition, so we do not think our results should be biased towards finding an increase in either ploidy level.

2.3.6 *Replicate evolution experiment*

Lastly, we re-evolved cultures maintained in the freezer to determine if we could recapitulate the original result of diploid takeover. We re-evolved culture revived from 1302 generations. We also initiated a second set of tubes where we spiked in a small number of 1488 generation diploids (i.e., the first time point after diploidy had swept) alongside the 1302 generation culture. We thawed completely the freezer tubes of the entire population that had been frozen after 1302 and 1488 generations during the initial evolution. 20ul aliquots from generation 1302 were inoculated into 10mL test tubes of YPD for the single time point replicates, while 18uL of culture isolated at 1302 generations was combined with 2uL of culture from 1488 generations for the mixed evolution replicates. Cultures were grown exactly as in the original experiment with 1:101 dilutions in 10mL YPD every day. In the first block of the experiment we re-restarted 20 test tubes from Gen1302 culture. The

initial (day0) tubes were maintained at 4°C for the duration of the experiment; after 26 days of transfers we transferred 100 μL of both day0 and day26 tubes into 10mL fresh YPD and allowed them to grow overnight. We then sampled the proportion of haploids and diploid from tubes using the same FACSCalibur protocol described above. A second block of the experiment was then initiated. We continued the initial tubes for 15 additional days. We also started a second set of evolution tubes; 10 new tubes were started from 1302 generation culture as well as 20 replicate tubes spiked with cells from 1488 generations. Statistical analyses was conducted to account for a block (or length of experiment) effect. We first compared the two blocks of evolution started from culture isolated at 1302 generations (41 days evolution vs. 15 days of evolution). We then compared the replicates started with 1302 generation culture against those started with 1302+1488 generation culture. A Fishers' exact test was used to test whether there was a significant increase in diploidy for each experiment. We tested for a difference between experiments (pure Gen1302 culture vs. Gen1302 + Gen1488) using a two-way t-test.

2.4 Results and Discussion

We sought to determine why diploids were able to overtake haploids during an ~1800 generation batch culture evolution experiment (Gerstein *et al.*, 2006). We have focused on one of five lines independently evolved in YPD ("Line A") that showed this pattern. By isolating many haploid and diploid colonies from throughout the initial evolution experiment we were able to assay fitness of colonies of different ploidy that were present in the same population. We previously found the ancestral strain to be aneuploid for chromosome IX (Gerstein *et al.*, 2008); though we have not tracked this aneuploidy directly in the experiments presented here. Variation for genome size was found by flow cytometry at several intermediate timepoints (Figure 2.1). We used the kmeans function in the R Programming Language (R Development Core Team, 2011) to assign each colony to a cluster, with the number of clusters (k) set to 2 (k=2 significantly decreased the within group sum of squares). As cluster assignments correspond to haploid and diploid genome sizes based on control samples, we refer to all colonies in the first cluster as haploids and colonies in the second cluster as diploids. The first diploid colony was sampled at 744 generations and the last haploid was sampled at 1302 generations, thus polymorphism for genome size was maintained for at least 558 generations.

We found consistent differences in cell size and shape between haploid and diploid colonies. Haploid cells isolated at both 1023 and 1302 generations had significantly smaller volumes and surface areas than diploid cells (Table 2.1, Figure 2.2), as is commonly observed (Weiss *et al.*, 1975; Mable, 2001). As predicted based on the equations for cell shape, diploid cells were also more eccentric (i.e., less spherical) and had a significantly lower surface area

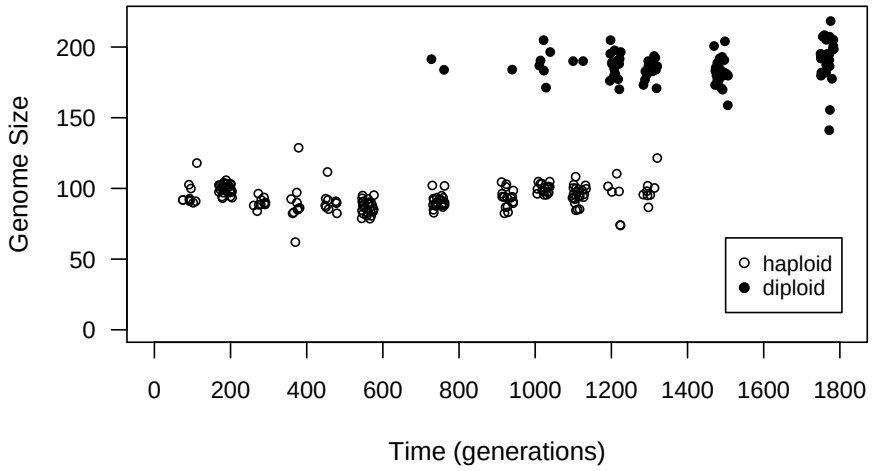


Figure 2.1: Polymorphism for genome size across the time series. 30 000 cells from each of 24 colonies were measured on a FACSCalibur at each time point, with haploid and diploid assignment determined by the kmeans function in the R programming language (R Development Core Team, 2011). Points are plotted with slight jitter on the x-axis for viewing purposes.

to volume ratio than haploid cells (Table 2.1, Figure 2.2; see also Mable 2001). Interestingly, we found evidence that cell size and shape may have changed within a ploidy level across the time frame of our initial experiment. As shown in Figure 2.3, we found a significant increase in both cell volume and surface area of haploid cells over time (using the lm function in the R programming language, R Development Core Team 2011); volume: $F_{1,18} = 5.87$, $p = 0.026$, surface area: $F_{1,18} = 6.48$, $p = 0.02$), with no significant change in eccentricity ($F_{1,18} = 0.35$, $p = 0.56$) or surface area:volume ratio ($F_{1,18} = 2.27$, $p = 0.15$). The only significant change for diploid colonies isolated at many time points was eccentricity; diploid cells became more elongated over time (eccentricity: $F_{1,15} = 5.58$, $p = 0.032$; volume: $F_{1,15} = 1.59$, $p = 0.23$; surface area: $F_{1,15} = 1.09$, $p = 0.31$; surface area:volume: $F_{1,15} = 1.78$, $p = 0.21$). An adaptive increase in cell size has previously been found for *E. coli* when evolved in minimal medium for 2000 generations under similar batch culture conditions (Mongold and Lenski, 1996). As previously mentioned, cell volume alone may contribute to potential differences between cells of differing ploidy (Wu *et al.*, 2010). Cell volume is highly correlated with surface area, eccentricity and surface area:volume whether we compare across the time series or within Gen1023 or Gen1302 colonies (statistics are presented in Tables A.1 and A.2). We thus focused only on cell volume as a potential correlate with growth phase components and fitness correlates of haploid and diploid colonies.

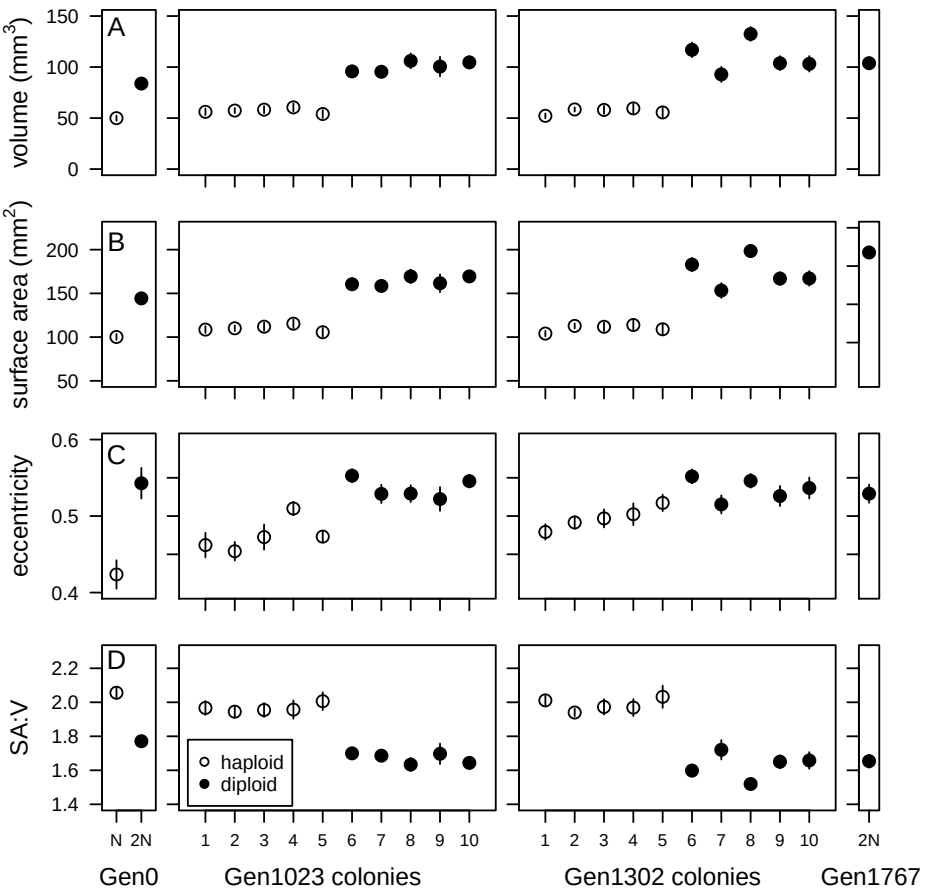


Figure 2.2: Cell size and shape at 1023 and 1302 generations. Cell size (A: volume, B: surface area) and shape (C: eccentricity, D: surface area to volume ratio) of ancestral and evolved populations, as well as five haploid and five diploid colonies isolated at 1023 and 1302 generations. Numbers assigned to a colony are used consistently throughout all assays (and in all figures).

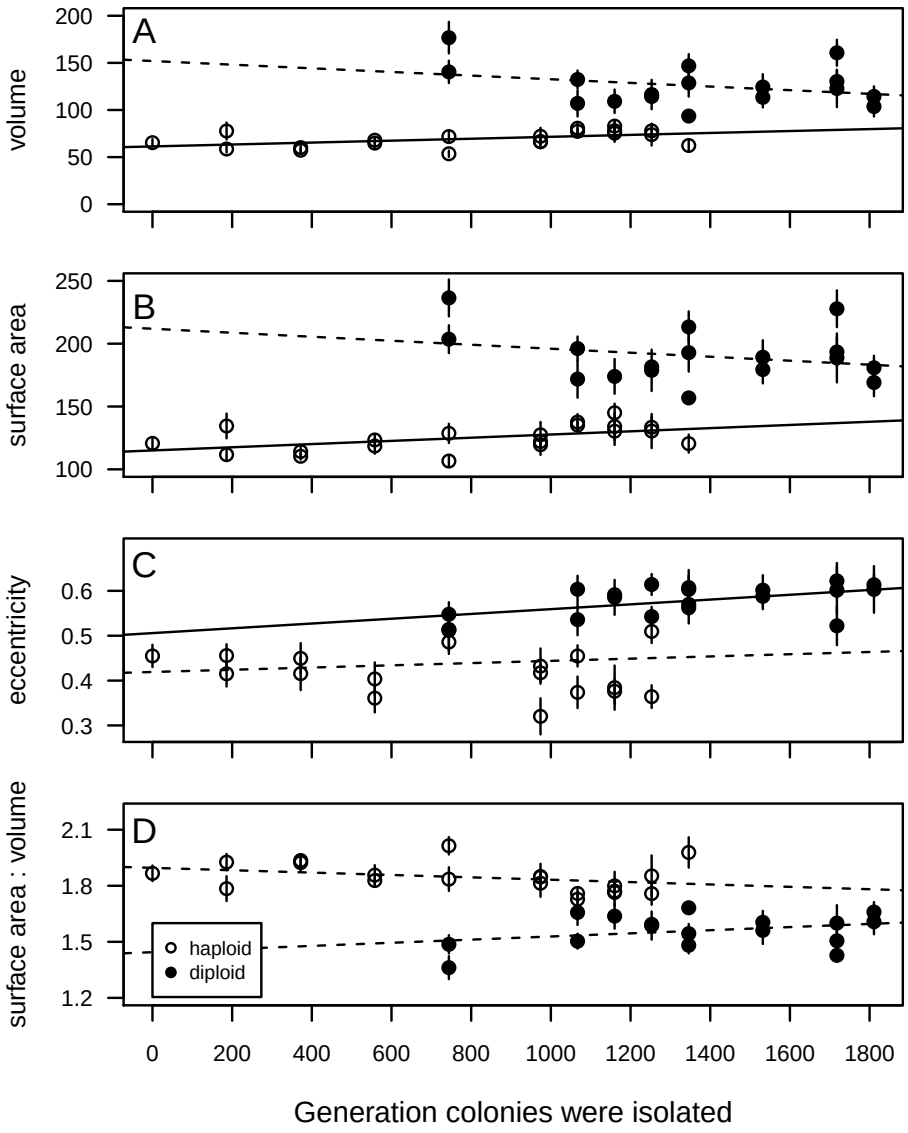


Figure 2.3: Cell size and shape across the time series. Cell size (A: volume, B: surface area) and shape (C: eccentricity, D: surface area to volume ratio) measures for 15 cells from two to four colonies isolated at each time point. Haploid and diploid colonies were analyzed separately to test the relationship between cell size/shape and generation of colony isolation. Solid lines indicate a a significant linear regression ($p<0.05$) while dashed lines indicate a nonsignificant trend ($p > 0.5$). Here and in later figures in this chapter, error bars represent $\pm$ 1 SE.

Table 2.1: Cell size & shape statistics comparing 20 haploid and 17 diploid colonies isolated across the full time series (Figure 2.3, haploid colonies isolated between 0 and 1346 generations, diploids between 744 and 1767 generations), and five haploid and five diploid colonies isolated at each of 1023 and 1302 generations of evolution (Figure 2.2). In each case we compared haploid and diploid colonies using a Welch two sample t-test, not assuming equal variance; *** p <0.0001, ** $p < 0.001$, * $p < 0.01$

	full time series	1023 generations	1302 generations
volume	$t_{20.3}$ = -10.21***	$t_{5.8}$ = -17.7***	$t_{4.3}$ = -7.6**
(N)	69.5 ± 1.4	57.3 ± 1.2	56.8 ± 1.5
(2N)	126.2 ± 3.2	100.5 ± 2.5	109.8 ± 7.6
surface area	$t_{22.4}$ = −11.8***	$t_{7.2}$ = -18.9***	$t_{4.4}$ = -8.0**
(N)	124.9 ± 3.2	110.3 ± 1.8	110.3 ±2.0
(2N)	190.8 ± 1.6	163.9 ± 2.6	173.8±8.7
eccentricity	$t_{28.0}$ = -7.5***	$t_{6.5}$ = -5.5*	$t_{8.0}$ = -4.1*
(N)	0.578 ± 0.008	0.449 ± 0.021	0.495 ± 0.014
(2N)	0.436 ± 0.008	0.572 ± 0.013	0.570 ± 0.015
surface area:volume	$t_{32.4}$ = 10.4***	$t_{7.5}$ = -16.8***	$t_{5.8}$ = -9.5***
(N)	1.84 ± 0.012	1.97 ± 0.012	1.99 ± 0.018
(2N)	1.56 ± 0.014	1.67 ± 0.016	1.63 ± 0.038

Three different growth phase components were measured in an attempt to capture the primary phases of growth experienced by yeast cells in the 24 hours between transfers during the original experiment (Gerstein *et al.*, 2006). Surprisingly, we found that none of these component fitness measures indicated an advantage to diploidy, despite diploids overtaking haploids. To test whether the lag phase of growth differed between haploids and diploids, we used mass spectrometry to measure the percentage of glucose remaining in the medium every 2 hours until 8 hours after transfer for four populations initiated from colonies isolated after 0 (haploids), 1302 (haploids and diploids), and 1767 generations (diploids). We also measured the percentage of glucose remaining in the medium at 24 hours. If lag phase differs between haploids and diploids, we expect to find differences in the glucose remaining at the early time points. As the amount of glucose present initially is the same, any difference in lag phase would be recovered as a difference in the amount of glucose remaining in the medium due to differences in the rate of glucose metabolism. However, as shown in Figure 2.4, a two-way ANOVA indicated that although the percentage of glucose decreases in the medium as post-transfer time increases ($F_{1,36}$ = 1359.8, $p <$ 0.0001), no differences were detected between haploid and diploid colonies ($F_{1,36}$ = 0.82, p = 0.37), nor was there a significant interaction between generation and ploidy ($F_{1,36}$ = 1.80, p = 0.19). Similarly, when we compared the four populations, ploidy did not significantly

affect glucose % over the first eight hours of growth (t-test; 2h: $t_{5.8}$ = 5.7, p = 0.7; 4h: $t_{8.0}$ = -1.5, p = 0.2; 6h: $t_{7.6}$ = -1.1, p = 0.3; 8h: $t_{7.3}$ = 0.69, p = 0.5). These results indicate that haploid and diploid colonies begin to grow and utilize glucose at similar rates, suggesting no difference in lag phase. Interestingly, after 24 hours diploid lines had significantly more glucose remaining in the medium than haploid lines ($t_{8.0}$=-5.4, p = 0.0007), suggesting that they are either less efficient at utilizing glucose for growth and biomass production or that they switch to metabolizing ethanol before glucose is used up.

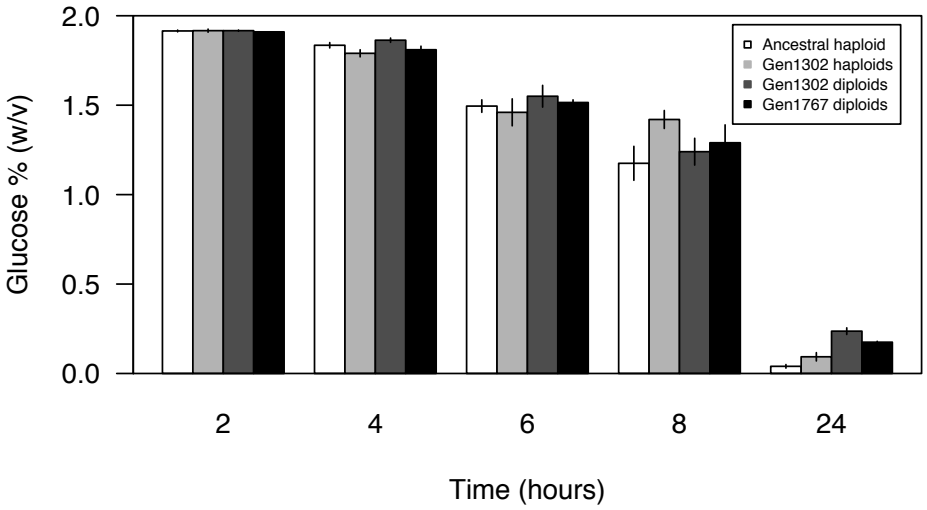


Figure 2.4: Lag phase fitness proxy. Glucose % (w/v) measured by HPLC post transfer into new medium. Haploids (isolated at the first time point and after 1302 generations of evolution) and diploids (1302 generation and 1767 generation) do not differ significantly in the amount of glucose present in the medium at the early time points, suggesting that their growth lags do not differ substatially.

Looking across the entire time series (Figure 2.5), neither growth rate nor biomass production predicted why diploid colonies might be able to invade haploids. Using a partial correlation test to remove the effect of time, haploid colonies both grew faster (t_{34} = -2.87, p = 0.004) and reached higher biomass (t_{34} = -5.46, $p < 0.0001$) than diploid colonies. The correlation between time and growth rate was not significant for either ploidy level (haploids: r = -0.14, t_{16} = -0.59, p = 0.57; diploids: r = 0.01, t_{14} = 0.05, p = 0.96) nor time and biomass production (haploids: r = 0.02, t_{16} = 0.07; p = 0.94, diploids: r = -0.41, t_{14} = -1.68, p = 0.11).

We then looked in greater depth at the populations from 1023 and 1302 generations. Diploid colonies had a lower growth rate than haploid colonies, significantly so at 1302 generations (Figure 2.6; Welch's two-sample t-test; 1023: $t_{6.8}$ = 0.33, p = 0.75; 1302: $t_{6.8}$ = 3.41, p = 0.01). Biomass production also did not differ significantly between haploid and diploid

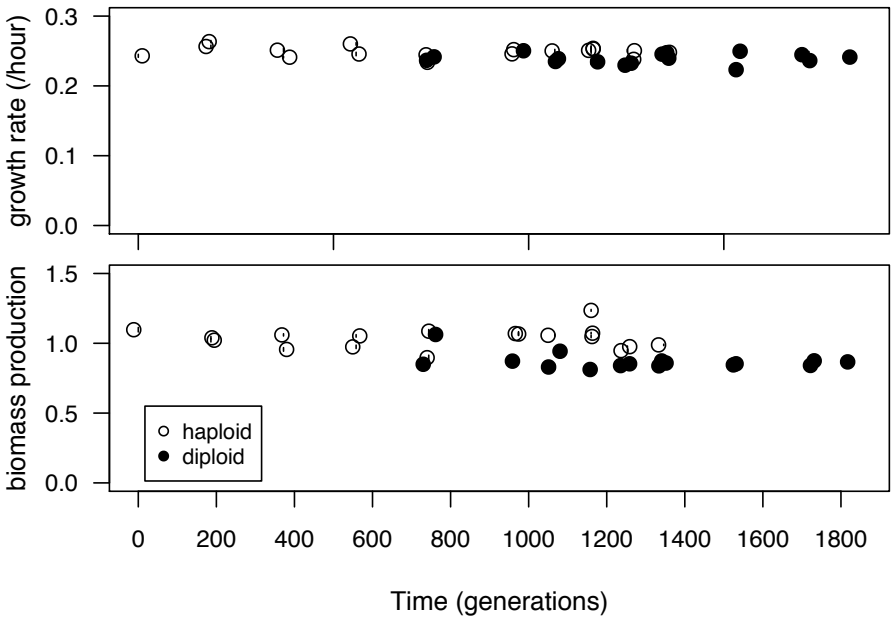


Figure 2.5: Growth rate and biomass production across the time series. Growth rate and biomass production were measured for 20 haploid and 17 diploid colonies isolated throughout the original experiment. Points are plotted with slight jitter on the x-axis for viewing purposes.

colonies when they were isolated at the same time point (1023 generations: $t_{7.0}$=-0.491, p = 0.64; 1302 generations: $t_{5.7}$=0.07, p = 0.95). Neither growth rate nor biomass production showed a significant correlation with cell volume for colonies isolated across the time series when we control for ploidy using a partial correlation (growth rate: p = 0.84; biomass production: p = 0.08) or when we examine colonies from 1023 and 1302 generations together (growth rate: p = 0.80; biomass production: p = 0.52). Although researchers typically assume that growth rate is the primary factor under selection in batch culture (Dykhuizen, 1990), our results do not support this. One caveat to this conclusion is that, because of the large number of colonies assayed in replicate, these parameters were measured in 100 well honeycomb plates rather than in the test tubes of the original experiment; it is possible that growth rate differences might have been apparent in a different environment.

When we measured population size after 24 hours (Figure 2.6), we found that diploid genotypes produce significantly fewer individuals within a growth cycle than haploids. Population size correlated very strongly with cell volume (t_{21}=-5.68, p <0.0001), though this relationship is driven entirely by ploidy, as there is not a significant correlation when we use a partial correlation to control for ploidy (p = 0.301) The fact that glucose consumption appears to be equal between haploid and diploid populations despite fewer individual diploid cells indicates that the average diploid individual metabolizes glucose faster than the average haploid individual. The differences we found in cell size and shape likely explain how diploid cells (which are fewer in number, but larger) are able to consume glucose at the same overall rate as haploid cells.

Selection may actually favour a slower growth rate if there is a tradeoff between growth rate and a second fitness component. For example, Blount *et al.* (2008) found that mutant *E. coli* that have acquired the ability to metabolize citrate outcompete individuals that cannot; citrate mutants have a significantly slower growth rate and a longer lag phase than other individuals isolated at the same time, yet they reach much higher optical density. Novak *et al.* (2006) explicitly tested for a tradeoff between growth rate and yield (biomass production) in 12 *E. coli* populations that had evolved for 20 000 generations. They did not find a significant tradeoff when comparing across the 12 populations with samples isolated at multiple time points. Interestingly, however, they do find evidence for significant tradeoffs when they look at many colonies isolated from the same population at one time point. We, however, do not find evidence for a negative tradeoff in our experiment, at least between growth rate and biomass production. When we compare across all timepoints (Figure 2.5), we find a significant positive correlation between growth rate and yield (t_{32} = 3.75, p = 0.0007, r = 0.55), yet this relationship is largely driven by the effect of ploidy. If we test for a correlation within haploid or diploid populations we find no significant correlation for either, but in both cases the correlation is positive (haploid: t_{16} = 0.83, p = 0.42, r = 0.20; diploid: t_{14} = 0.95, p = 0.36, r = 0.25). Our findings are similar to results obtained by Adams

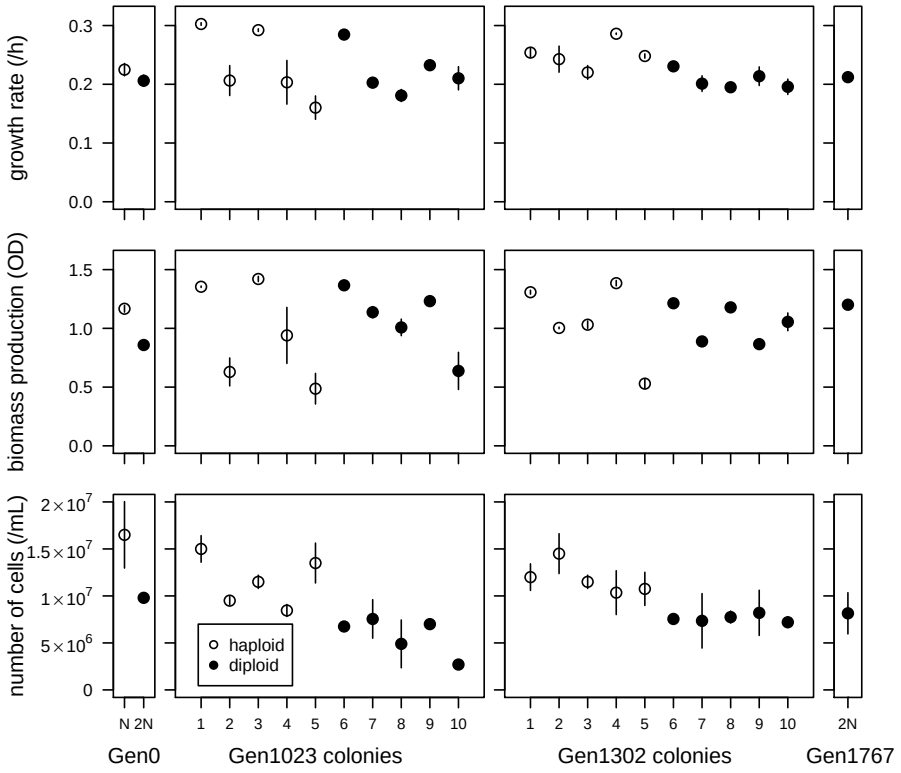


Figure 2.6: Growth rate, biomass production and the number of cells at 24 hours. Growth rates (top), biomass production (middle) and the density of cells (bottom) measured for ancestral haploid and diploid lines, five colonies of each ploidy after 1023 generations and 1302 generations of evolution, and the diploid population after 1767 generations.

et al. (1985) when examining *S. cerevisiae* clones isolated during 260 generations of growth in a chemostat. Specifically, Adams *et al.* found that growth rates changed very little over the course of the experiment and that the growth rate of some of their later time point clones is lower than the initial clones. It thus seems that neither growth rate nor biomass production are able to explain why diploids overtook haploids in our evolution experiment.

Fitness assays that examine population level parameters may not capture the dynamics that occur when different genotypes (or cells of differing ploidy) are in direct competition with each other, either because of interactions between individuals of different ploidy levels or because of unmeasured components of fitness in batch culture. For example, previous work on *Candida albicans* did not find a significant correlation between fitness measured by direct competition experiments and fitness measured on isolated populations (examining either growth rate or stationary phase density; Cowen *et al.* 2001). We thus turned our attention to competition assays that account for interactions between different ploidy types and that integrate fitness across the entire 24 hour batch culture cycle.

We first determined the competitive ability of generation zero haploid colonies and generation 1767 diploid colonies, as well as haploid and diploid colonies isolated at 1023 and 1302 generations, against a closely-related marked competitor (both our ancestor and the common competitor are derivatives of *S. cerevisiae* strain S288C). This assay was, however, also unable to explain why diploids were able to overtake haploid colonies (Figure 2.7). Altogether, we found that only the generation at which a colony was isolated significantly affected competitive ability (two-way ANOVA, time: $F_{1,19} = 22.9$, $p = 0.00013$, ploidy: $F_{1,19} = 3.18$, $p = 0.091$, time*ploidy: $F_{1,19} = 3.21$, $p = 0.089$). When we look at the difference in competitive ability between colonies of different ploidy isolated at 1302 generations, we find that haploids compete significantly better than diploid colonies (two-way t-test: $t_{4.5}=2.88$, $p = 0.039$), although ploidy did not significantly affect competitive ability among the 1023 generation colonies (two-way t-test: $t_{5.1} = 0.458$, $p = 0.666$).

Experiments that compete colonies against a common competitor (or the ancestor) also do not precisely mimic the original evolution experiment, however. If non-transitive fitness changes are occurring, comparing fitness against the ancestral type does not inform us about competitive ability against the actual genotypes that were present at any point in time. Such non-transitive fitness interactions have been shown to be important in some previous microbial experiments (Paquin and Adams, 1983), but not others (de Visser and Lenski, 2002). To control for this potentially important factor, the best fitness assay is one that directly competes colonies from the same time point together. We thus competed a population of 6 haploid colonies isolated at 1302 generations against single diploid colonies isolated at 1302 generation (for 12 individual diploid colonies), as well as against a pooled populations of 12 diploids colonies isolated from 1302 generations ("Gen1302 2N population") and a pool of 24 diploid colonies isolated at 1488 generations, after diploids appeared

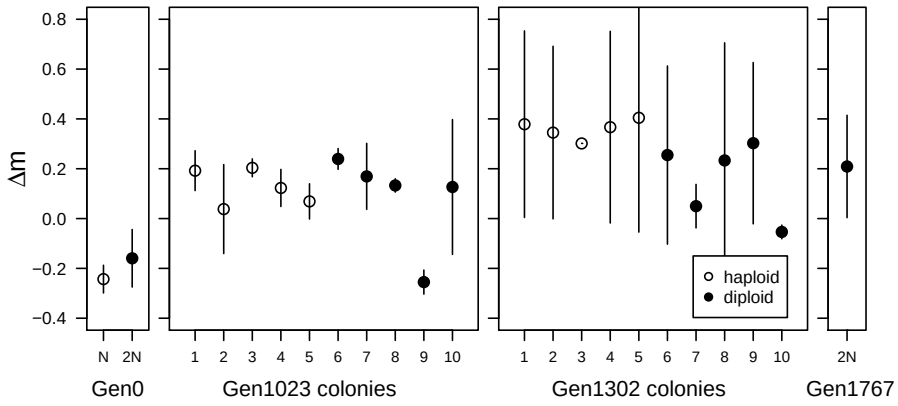


Figure 2.7: Haploid and diploid colonies were competed directly again a common marked competitor for 72 hours. The y-axis (Δm) is the difference in malthusian growth rate between the given strain and the common competitor.

to have fixed ("Gen1488 2N population"). As we are primarily interested in the ability of diploids to overtake haploids (as was observed in the original experiment), we conducted a one-way t-test to look for a significant increase in diploid frequency after 14 days of competition. As shown in Figure 2.8 & Table 2.2, only the population of diploids isolated at 1488 generations, i.e., the first generation where only diploid colonies were sampled, were significantly able to increase in diploid frequency. None of the single diploid colonies nor the population of diploids created by combining single colonies from 1302 generations significantly increased in frequency compared to the haploid population from 1302 generations (Table 2.2; one of the previously assayed diploid lines became visibly contaminated during the experiment and was not assayed, "colony 4" in the first panel of Figure 2.8).

In summary, none of our fitness assays predicts that diploid colonies isolated from either 1023 or 1302 generations would be able to overtake contemporaneous haploids present in the environment at the same time point during the original experiment. Rather, our results indicated that the fitness advantage that allowed eventual takeover by diploids arose in only a subset of diploid lineages, which predominated by 1488 generations. Recall, however, that growth rates were not significantly higher at the end of the experiment (Figure 2.6, Gen1767), nor were they significantly higher at generation 1488 (Figure A.4), leading us to conclude that growth rate measures fail to predict competitive advantage in these diploids.

To further explore the population dynamics of the population isolated at 1302 generations, we re-evolved freezer culture that was acquired during the original experiment. Culture isolated after 1302 generations of evolution was re-evolved during two blocks, one that lasted for 41 days and one that lasted 15. There was no significant difference in the total

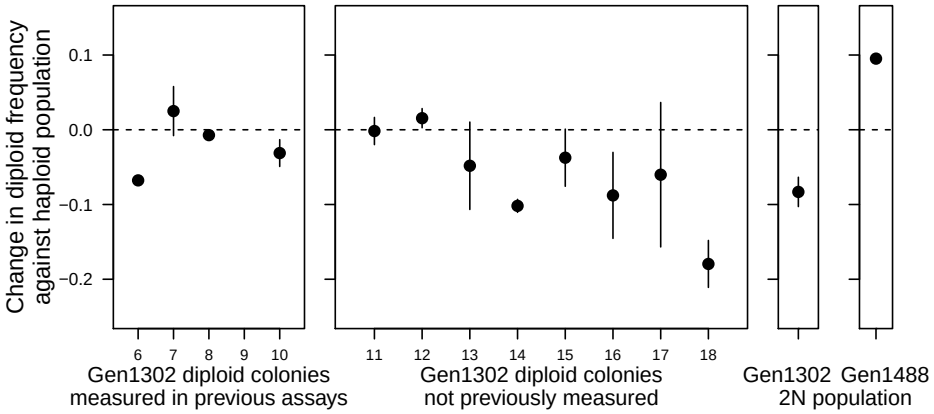


Figure 2.8: Competition against the haploid population from 1302 generations. Thirteen diploid colonies isolated at 1302 generations, a pool of 12 1302 generation diploids ("Gen1302 2N population") and a pool of 24 diploid colonies isolated at 1488 generations ("Gen1488 2N population") were competed against a population of the 6 haploid colonies isolated at 1302 generations. Contamination arose in one of the diploid colony competitions, and we were unable to measure the results of this competition (the blank space in the first panel). Only the population of diploids from 1488 generations (the first time point after diploidy swept in the original experiment, rightmost panel) was consistently able to outcompete the haploid population. All competitions were started at 50:50 (v/v) with transfers into fresh medium every 24 hours for 14 days. Standard error bars based on two replicate competitions.

Table 2.2: T-test results of single diploid colonies isolated from 1302 generations and diploid populations from 1302 and 1488 generations competed directly against a population of haploids isolated from 1302 generations. Colony ordering as in Figure 2.8; the first 5 colonies are the same five colonies measured in the other fitness experiments. The assay compares the frequency of diploid cells after 14 days of competition using a FACSCalibur.

competitor	one-way t-test
1302gen - colony 6	t_1 = -12.3, p = 0.97
1302gen - colony 7	t_1 = 0.8, p = 0.29
1302gen - colony 8	t_1 = -1.14, p = 0.77
1302gen - colony 9	contaminated
1302gen - colony 10	t_1 = -1.7, p = 0.83
1302gen - colony 11	t_1 = -0.1, p = 0.53
1302gen - colony 12	t_1 = 1.21, p = 0.22
1302gen - colony 13	t_1 = -0.8, p = 0.72
1302gen - colony 14	t_1 = -11.8, p = 0.97
1302gen - colony 15	t_1 = -1.0, p = 0.75
1302gen - colony 16	t_1 = -1.5, p = 0.82
1302gen - colony 17	t_1 = -0.6, p = 0.68
1302gen - colony 18	t_1 = -5.7, p = 0.94
1302gen - 2N population	t_1 = -4.2, p = 0.93
1488gen - 2N population	t_1 = 15.0, p = 0.02

change in diploid frequency between blocks ($t_{14.1}$ = -0.35, p = 0.73), thus we combine them for analysis. Across 30 replicate tubes initiated from a population sample taken from 1302 generation culture, there was no significant change in diploid frequency (t_{29}=-0.94, p = 0.35), with 14 tubes showing an increase in diploid frequency and 16 tubes showing a decrease (Figure 2.9A), with tremendous variation among replicates. The set of replicates initiated with 18 μL culture isolated after 1302 generations spiked with 2 μL culture from 1488 generations more often exhibited an increase in diploid frequency, but the change in diploid frequency was again not significantly different than 0 (Figure 2.9B: 13 test tubes increased in diploid frequency, 6 decreased; t_{18} = 1.4, p = 0.17). The starting diploid frequency was not significantly different between the two treatments ($t_{46.1}$ = -0.127, p = 0.90), yet diploid frequency did increase significantly more when 1302 generation culture was spiked with 5% 1488 generation culture than when it was not (one-way Welch t-test: $t_{37.0}$=-1.70, p = 0.049).

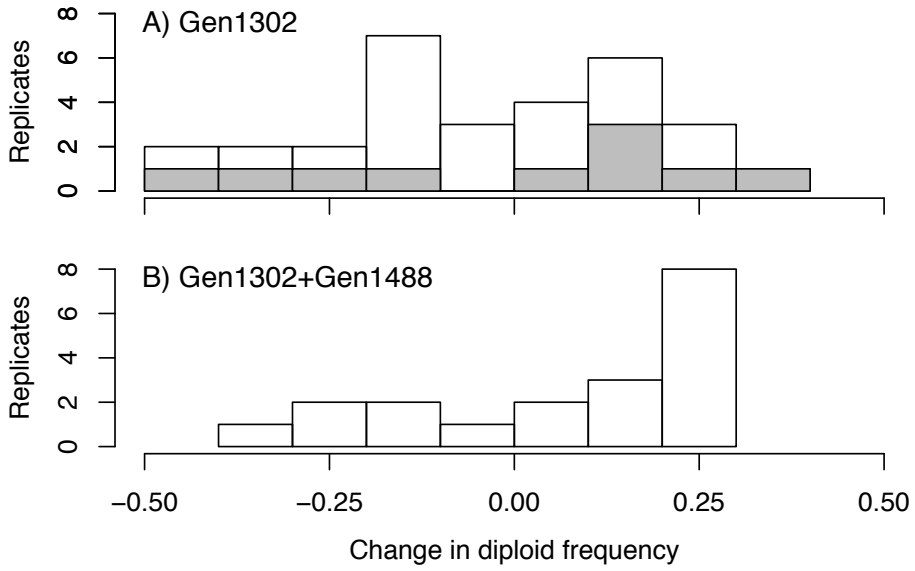


Figure 2.9: Replicate evolution experiment. The experimental evolution study was restarted from A: culture from 1302 generations and B: culture from 1302 generations spiked with 5% diploid culture from 1488 generations. Replicates were evolved through batch culture in exactly the same way as the original experiment (Gerstein *et al.*, 2006). The 1302 generation culture was evolved for 41 days (20 replicates) and 14 days (10 replicates - shaded bars in top panel), mixed culture was evolved for 14 days (20 replicates).

2.5 Conclusions

The experiments described above aimed to determine how diploid individuals were able to rise in frequency within the ancestral population of haploids. Our results failed to find any fitness advantage of early-arising diploids (generations 1023 and 1302) over haploids. How could the diploids have risen to intermediate frequency without a fitness advantage? Several possible explanations remain. The environments might have been slightly different than in the initial experiments. Alternatively, even though we sampled five diploid colonies at both of these two time points, perhaps we were unlucky and sampled particularly unfit diploids. More likely, we note that none of these standard fitness assays would have revealed a fitness advantage if such an advantage is negative frequency dependent. Competition assays starting at different initial frequencies of diploids suggests that the diploids from generations 1023 and 1302 are able to spread when rare, but this competitive advantage declines with frequency (A.C.G., in prep). We hypothesize that this is why the standard fitness measures used here failed to explain the initial rise in diploid frequency.

Our results indicate that there is not a fitness-related trait that uniformly differs between haploids and diploids and that allows diploids to overtake haploids whenever they appear. Rather, we conclude that only a subset of diploids, which predominated late in the experiment (generation 1488), are competitively superior and capable of fixing within the population (Figure 2.6). Our direct competition assay and replicate evolution experiment both suggest that the diploids we sampled (specifically, diploids colonies isolated at 1302 generations) are unlikely to be the same exact diploid genotypes that overtook the haploid population during the initial experiment. Interestingly, fitness measures from later generations failed to show evidence of higher growth rates or biomass production (Figures 2.4, A.1), although diploids from these later generations exhibit competitive superiority (Figures 2.8 and 2.9). Current sequencing efforts aim to identify the causative mutations underlying the advantage of later generation diploids. We could then determine whether the mutation was accessible or beneficial to diploids alone, explaining the consistent conversion of haploid to diploid populations.

The exact selective forces acting within our experiment remain largely unknown. It may be that organisms are adapting to an aspect of the medium (YPD), to the test tube environment (e.g., low oxygen), or to batch culture (i.e., repeated feast and famine). One clue, however, might be that the smaller haploid cells significantly increased in size (approximately 1% increase in volume per generation). An adaptive increase in cell size has also been found for *E. coli* when evolved in minimal medium under similar batch culture conditions (Mongold and Lenski, 1996). We hope that future experiments and sequencing efforts will help shed light on this question.

We are left to conclude that the evolutionary dynamics of this system are more complicated than expected, and that none of the standard assays used to measure fitness demonstrate diploid superiority over haploidy across all diploid lines. The picture that emerges is that the ploidy level of any given colony isolated from a particular time point is not the determining factor in whether that individual has high fitness and will spread. We find tremendous trait variation among colonies of the same ploidy level for the majority of traits measured, and the variation among colonies of the same ploidy is often larger than the variation between ploidy levels (e.g., Figure 2.5). If anything, haploid cells appear to have the higher fitness (for growth rate and biomass) at intermediate time points when both haploids and diploids are present. As Adams *et al.* (1985) noted at the end of their paper examining a chemostat-evolved population of *S. cerevisiae* twenty five years ago:

> *"The emerging picture of adaptation in such populations, therefore, is that a number of different cell phenotypes may exhibit increased fitness and that the selection of any one of them is unpredictable, depending on the random nature of the mutational events involved. [We believe] a single optimal phenotype may not exist even for simple constant laboratory environments."*

Although the role of ploidy in our previous evolution experiment (Gerstein *et al.*, 2006) seems to be deterministic in that diploids eventually outcompeted haploids in all ten of our replicate lines, ploidy is not the most important differentiating character among cells present in the population. These experiments demonstrate the utility of maintaining a fossil record during batch culture evolution, allowing us to reconstruct the history of selection.

Chapter 3

Haploids adapt faster than diploids across a range of environments

3.1 Summary

Despite a great deal of theoretical attention, we have limited empirical data about how ploidy influences the rate of adaptation. We evolved isogenic haploid and diploid populations of *Saccharomyces cerevisiae* for 200 generations in seven different environments. We measured the competitive fitness of all ancestral and evolved lines against a common competitor and find that in all seven environments haploid lines adapted faster than diploids, significantly so in three environments. We use theory to show that there are likely only single mutations at high frequency in any given line and find that across 35 haploid lines evolved in seven environments the effect size of selected mutations varied from 0-25%. By comparing the rate of adaptation between haploid and diploid populations we find evidence that these first selected mutations are likely to be additive to dominant in fitness. These results are consistent with theory that predicts haploids should evolve faster than diploids at large population sizes.

3.2 Introduction

The rate at which beneficial mutations arise and fix determines how quickly a population can adapt to novel environments. This is particularly important for populations in very stressful environments, where to avoid extinction, novel beneficial alleles must spread fast enough to counter fitness declines due to external environmental change and internal accumulation of deleterious alleles (Orr and Uncкless, 2008; Bell and Collins, 2008). The rate

of adaptation is affected by various properties that determine the fixation rate of beneficial alleles: the availability of mutations (mutational neighbourhood, Burch and Chao 2000), the mutation rate (μ), the distribution of fitness effects (s), and the dominance of mutant alleles (h). Here we compare the rate of adaptation between haploid and diploid initially isogenic lines of *Saccharomyces cerevisiae* in seven different environments. Comparing rates of adaptation between ploidy levels and across many environments allows us to make inferences about the the genetic properties of the mutations contributing to adaptation in these experiments.

The effect of ploidy on ecology and evolution has long been a question of interest (Adams and Hansche, 1974, and references within). Isogenic haploid and diploid populations of *S. cerevisiae* allow direct comparisons between individuals that share a genotype but differ in ploidy. Even with identical genomes, ploidy itself is known to have several direct effects on yeast. A recent study found that 2.7% of the proteome changed more than 50% in abundance between isogenic haploid and diploid cells (de Godoy *et al.*, 2008). Proteins that differed were in the pheromone pathway (specific to haploid cells), retrotransposon-associated proteins (ten times more abundant in haploids), and cell wall components, which were downregulated by a factor of 0.77 in diploids. Interestingly, this level of reduction in cell wall components is close to that predicted (0.79) based on geometric considerations. By nature of cell geometry, the surface area to volume ratio decreases as cell size increases, and thus haploids, which are smaller than diploids, have a greater proportional surface area than diploids cells (Weiss *et al.*, 1975). The difference in cell size between haploids and diploids is predicted to directly affect their relative fitness in some environments. Under nutrient stress, for example, where the limiting nutrient diffuses across the cell membrane, haploids are expected to have an advantage (Weiss *et al.*, 1975; Lewis Jr, 1985). Support for this hypothesis has been found in some studies (Adams and Hansche, 1974) but not others (Mable, 2001).

At present, we have only fragmentary knowledge about how the frequency and properties of novel beneficial mutations are affected by ploidy levels. Some evidence suggests that the availability of particular types of beneficial mutations may differ by ploidy. Gresham *et al.* (2008) found that diploids were more likely than haploids to select large amplification and deletion mutations during a 200 generation chemostat experiment. Similarly, Thompson *et al.* (2006) found a class of mutations selected among diploid mutator strains that conferred an advantage across a range of environmental conditions; these mutations did not appear in haploid mutator strains or in non-mutator lines of either ploidy. Further investigation revealed that the specific mutation may have been a chromosomal rearrangement, which was potentially beneficial to heterozygous diploids but deleterious or neutral to haploid cells. The mutation rate could also differ between haploids and diploids. Though one experiment found that the mutation rate per base pair was the same (Ohnishi *et al.*,

2004, 1.06×10^{-6}), a second experiment found that microsatellite stability in the mitochondrial genome was not (Sia *et al.*, 2003, found a 100-fold higher mutation rate in haploids). Paquin and Adams (1983) show that diploids adapt faster than haploids by estimating the frequency of adaptive mutations in five haploid and six diploid lines evolved for up to 300 generations in glucose-limited chemostats, although it has been argued that the fluctuation assay identified a larger number of mutations than could have fixed during the course of their experiments (Dykhuizen, 1990).

Regardless of mutation rate and availability, the efficacy of selection is predicted to differ between haploids and diploids. The fixation probability of a beneficial mutation in a diploid is approximately $2hs$ (Haldane, 1927), where h denotes the dominance of the mutation, i.e., how much of its homozygous fitness benefit (s) is experienced in a heterozygote. If beneficial mutations are on average recessive (i.e., a single mutated allele is masked by the wildtype, $h < 0.5$), diploids should adapt at slower rates than haploids despite having twice the number of mutational targets (Orr and Otto, 1994). If beneficial mutations are sufficiently dominant and mutations are limiting, diploids, with double the mutational targets, can evolve faster (Orr and Otto, 1994). These theoretical predications require that the population size and selective effects (s) are equal between haploids and diploids. The dominance and availability of beneficial mutations are thus both expected to influence whether haploids or diploids adapt faster in a particular environment.

The distribution of dominance of beneficial mutations remains unknown, though empirical results have shown that the dominance of mutations does significantly affect the relative rate of adaptation of haploids and diploids. Zeyl *et al.* (2003) evolved haploid and diploid *S. cerevisiae* populations asexually for 2000 generations in minimal medium at large population sizes (where selection was the primary evolutionary force acting). They found that haploid populations adapted significantly faster than the diploids, and consistent with theory (Orr and Otto, 1994), the average dominance of the beneficial mutations selected in one of the haploid lines was 0.20. Anderson *et al.* (2004) also demonstrated the potentially critical role of dominance in determining the relative rates of adaptation by adapting *S. cerevisiae* to the drug fluconazole. At low concentrations of fluconazole, resistance is primarily achieved through dominant mutations in the *PDR1* gene, while predominantly recessive mutations in *ERG3* are fixed at high concentrations (Anderson *et al.*, 2004). When haploid and diploid populations were evolved for 100 generations to low concentrations of fluconazole, diploids, with double the mutational targets, evolved faster. In contrast, at high concentrations of fluconazole, haploids were able to fix the required recessive mutations and adapted faster than diploids.

In this paper we have compared the rate of adaptation between haploids and diploids evolved at large population sizes in seven different environments for 200 generations. We find the broad pattern to be identical across environments – haploids adapted faster than

diploids – though the magnitude of this difference varied across environments. We measured haploid and diploid effective population size (N_e) and found that haploid populations are significantly larger than diploids in almost all environments; previous theory that assumed equal population sizes was thus adjusted to allow this parameter to differ between ploidy populations. Combining these results, we use theory to predict the average dominance of the first beneficial mutations selected during these adaptive walks and find them to be consistently semidominant to dominant. Our results indicate that haploid microbes are likely to evolve faster than diploids across an array of environmental challenges.

3.3 Methods

3.3.1 *Haploid & Diploid Lines*

Initially isogenic haploids and diploids of haplotype MATa-a1 ste6△8-694 ura3 leu2 his4 trp1 can1 were created as previously described (Mable and Otto, 2001). Results obtained after the experiments reported here were initiated showed that both the haploid and diploid ancestral clones are aneuploid for chromosome 9 (haploid one extra copy, diploid two extra copies, Gerstein *et al.* 2008). This is not expected to affect our results strongly, as chromosome 9 aneuploids have similar cell volumes and doubling times as wildtype (Torres *et al.*, 2007).

3.3.2 *Environments & Experimental Evolution*

In addition to a standard rich medium (YPD), six stressful environments were used in these experiments. Moderately high levels of the following stressors were added to YPD: ethanol, salt (NaCl), caffeine, nystatin, potassium hydroxide (KOH), and hydrochloric acid (HCl). A brief description of the stressful environments and their major cell targets is provided in Table 3.1, and more complete methods are provided in Appendix B. These stressors affect yeast cells in a variety of ways, though (as with many stressors of single-celled organisms) all affect some aspect of the cell membrane or cell wall, which, as noted above, represents a primary phenotypic difference between isogenic haploids and diploids. The specific stressors were not chosen with any a priori expectation or prediction about their differential effect on haploids or diploids. The level of each stressor was chosen so that the initial growth rate was reduced by $\sim$ 20% relative to that in YPD of both haploids and diploids (unpublished results).

The isogenic haploid and diploid cultures were streaked from freezer stocks maintained at -80°C and grown on YPD plates for 48 hours. A single colony for each ploidy level was picked randomly and grown for 48 hours in 10mL YPD. Each ancestral ploidy culture was

Table 3.1: Evolutionary environments used to compare rate of adaptation

Stressor	Stress Level	Major Cell Targets
Caffeine	4.23mM	Mutagen (acts as a purine analog)[1]; inhibits repair of double strand breaks[2] and/or overrides DNA damage checkpoints[3]; affects metabolite transport across the cell membrane and protein translocation[3].
Ethanol	4%	Increases membrane fluidity; inhibits glycolytic enzymes; increases protein denaturation; affects transport systems such as general amino acid permease and glucose uptake; induces mutations in mitochondrial DNA[4].
HCl	pH $\sim$2.8*	Initiates yeast general stress response pathway; increases ROS production[5]; induces HOG-1 dependent cell wall organization changes[6].
KOH	pH $\sim$7.6*	Disrupts membrane proton gradients and uptake of solutes from the medium[7]; decreases nutrient and ion limitation[8]; can lead to cell wall damage; source of oxidative stress[8].
NaCl	0.6M	Decreases cell volume and turgor pressure[9]; causes hyperosmotic and ionic stress[10]; can decrease ATP hydrolysis[11].
Nystatin	0.6μM	Fungicide that causes membrane leakage[12]; increases permeability to protons[13]; alters vacuolar membrane and vacuolar morphology[14].

* A constant amount of HCl and KOH was added to YPD each time new medium was autoclaved; because of minor variation in autoclave protocol (e.g., how long bottles remained in the autoclave) the pH varied slightly for each medium batch.

[1]Kuranda *et al.* 2006, [2]Hannan and Nasim 1977, [3]Blasina *et al.* 1999, [4]Aguilera *et al.* 2006, [5]Giannattasio *et al.* 2005, [6]Kapteyn *et al.* 2001, [7]Lamb *et al.* 2001, [8]Serrano *et al.* 2006, [9]Nevoigt and Stahl 1997, [10]Matsumoto *et al.* 2002, [11]Nass *et al.* 1997, [12]Bard *et al.* 1980, [13]Palacios and Serrano 1978, [14]Bhiyan *et al.* 1999

used to inoculate five replicate lines in the seven different environments (six stressors plus a YPD control) for a total of 70 lines (2 ploidy $\times$ 5 replicates $\times$ 7 environments). 100μL stationary phase culture was transferred into 10mL fresh medium (i.e., 1:101 dilution) every 24 hours ($\pm$1 hour) for all 70 lines; under this regime there are approximately 6.7 generations of evolution between transfers ($2^{6.67} \approx 101$). Cultures were maintained at 30°C with continual shaking (200 rpm). Each line was evolved for a total of $\sim$189 generations with aliquots taken and frozen at -80°C in 15% glycerol every 49 generations. The length of the experiment was chosen because previous experiments had demonstrated that these haploid lines tend to diploidize under stressful conditions over longer time periods (Gerstein *et al.*, 2006). The ploidy of all evolved lines at four time points (49, 98, 147 & 189 generations) was checked at the conclusion of the experiment using flow cytometry (methods described in Gerstein *et al.* 2006), and no changes were observed.

Contamination by other microorganisms was checked under a microscope for all cultures every 24 hours (i.e., every seven generations). In addition, culture was periodically plated onto synthetic-complete plates lacking leucine; any growth on these plates indicated a contaminant (or possibly a revertant). Thirteen different experimental lines did show contamination at different points during the experiment with a variety of other microorganisms, and in each case we returned to the tube prior to the contamination and restarted the experiment from that time point (all evolved cultures were kept in tubes at 4°C for $\sim$ 4 days). The contaminants detected appeared haphazardly among lines. Although cross-contamination among lines within our experiment could not be detected by these methods, the fact that changes in ploidy level, which would be expected in 50% of cross-contamination events, were not observed by flow cytometry suggests that cross-contamination was absent or rare.

3.3.3 *Measuring competitive fitness*

Competitive fitness against a reference strain was used as a proxy for total fitness. The reference strain was constructed from BY74741 (MATa his3$\triangle$1 leu2$\triangle$0 met15$\triangle$0 ura3$\triangle$0) obtained from Open Biosystems. We insert a 3320-bp region of the pJHK043 plasmid containing YFP (yellow fluorescent protein) under control of the ACT1 promoter linked to a histidine marker, generously provided by John Koschwanez (FAS Center for Systems Biology, Harvard University). The region was isolated and amplified with primers F1 (TTCTTCGAAGAATATACTAAAAAATGAGCAGGCAAGATAAACGAAGGCAAAGATGCGTACG
CTGCAGGTCGACGG) and R1 (TACACATGTATATATATCGTATGCTGCAGCTTTAAATA
ATCGGTGTCACTACATACAGATCCGCGGCCGCATAGG) following J. Koschwanez (pers. comm.). This cassette was then inserted into BY4741 at the HIS locus and successfully transformed cells were selected on -his plates.

To quantify the rate of adaptation we determined the early (generation 49) and late (generation 189) competitive fitness for all 70 strains in their evolutionary environments. To eliminate any potential differences due to acclimation (i.e., non-genetic changes) between haploids and diploids we allowed the cultures to grow for an initial phase (0 to 49 generations) before tracking fitness. Competition assays were done separately for each of the seven environments; each competition assay involved 80 tubes (2 ploidy $\times$ 2 time points $\times$ 5 lines $\times$ 4 replicate competitions). The reference strain and competing strains were streaked onto YPD plates from freezer stocks maintained at -80°C and grown for 48 hours at 30°C. Culture was then inoculated from plates into tubes containing 10mL of their experimental environment (YPD for the reference strain) and grown overnight at 30°C, shaken at 200 rpm.

All competition assays except nystatin were initiated by inoculating 50μL from both the reference and competing strains into 10mL of the experimental environment. The nystatin competitions were initiated with 75μL reference and 25μL competing strains (see Appendix B for justification). Transfers were performed every 24 hours for 5 days in a manner that exactly mimicked the evolution experiment (100μL stationary phase culture was transferred into 10mL fresh medium with growth maintained at 30°C with continual shaking at 200 rpm). The ratio of fluorescing to non-fluorescing cells was measured on days 0, 2 and 4 after initiation (days 0, 2, 3 and 4 for nystatin). On each measurement day we placed 300μL aliquots into a 96 well plate exactly two hours after transfer. Plates were spun down for 3 minutes at 2500rpm. The supernatant was removed, and pellets were re-suspended in 300μL sodium citrate.

96 well plates were read on an LSRII flow cytometer with the High Throughput Sampler attachment. 10000 cells were measured for each well. The raw data (.fcs files) were export into FlowJo version 8.7. An initial gate was set by looking at the forward scatter (FSC-W) and side scatter (SSC-W) data to exclude small debris; this gate included between 95-99% of total events recorded. The data were plotted on FITC-A (x-axis) and AmCyan-A (y-axis), which provided maximal separation of fluorescing and non-fluorescing cells. Gates were drawn around the two distinct clusters of non-fluorescing and fluorescing cells (Figure B.1). All gates were set at the beginning of the experiment and were not subsequently altered.

For each line of interest we thus have four replicate competitions at three time points (day 0, 2 and 4 of competition, which correspond to 0, 13.4 and 26.8 generations. The competitive fitness (m) was determined for each line using the formula for evolutionary change:

$$NonFluor = \frac{p_0 e^{mT}}{1 - p_0 + p_0 e^{mT}} \tag{3.1}$$

where *NonFluor* is the fraction of non-fluroescing cells, p_0 is the initial fraction of non-fluorescing cells at the start of the experiment, T is the generation number and m is the Malthusian parameter of the experimental strain minus that for the YFP-marked competitor (relative growth rate). We use the nls function in the R programming language (R Development Core Team, 2008) to determine the best fitting p_0 and m for each competition. We measured the rate of adaptation as the rate of change in competitive fitness (m) for each of the 70 strains evolved in this experiment by calculating the slope over time (from generation 49 and 189) of the best fitting linear model using the lm function in R (R Development Core Team, 2008). We compared the five haploid slopes to the five diploid slopes in each environment using a two-sample t-test with the Welch modification for degrees of freedom, which does does not assume equal variance between groups.

3.3.4 *Effective population sizes*

The number of cells produced after 24 hours of growth for all ancestral (generation 49) and evolved (generation 189) lines was determined by a plating experiment. All lines were streaked onto YPD plates from freezer stocks maintained at -80°C and grown for 48 hours at 30°C. Culture was inoculated from plates into tubes containing 10mL of their experimental environment and grown overnight at 30°C, shaken at 200 rpm. We then mimicked the evolution experiment exactly by transferring 100 μL overnight culture into fresh medium and allowing cells to grow for exactly 24 hours. After 24 hours we diluted culture and plated three different dilutions onto 3 plates each.

We use theory developed by Campos *et al.* (2008) to calculate the effective population size with periodic bottlenecks as $N_e = r^2 \tau N_0$, where τ equals the number of generations between bottlenecks (6.7 in our experiment) and r is the growth rate. We can use the equation $N_f = N_0 e^{r\tau}$ to isolate r as $\frac{ln\frac{N_f}{N_0}}{\tau}$. Since $\frac{N_f}{N_0}$ equals the dilution rate (101 in our experiment) and τ is 6.7, r equals 0.689 and we simply multiply our final number of cells (N_f) by 0.031 to obtain the effective population size (N_e).

3.4 Results

Haploids were found to adapt faster than diploids in all seven environments (Figure 3.1). We first conducted a two-way ANOVA comparing all haploid and diploid slopes (change in competitive fitness) across all environments. There was a significant difference between ploidy levels ($F_1 = 24.7$, $p < 0.0001$) and across environments ($F_6 = 18.95$, $p < 0.00001$) with no significant interaction ($F_6 = 0.711$, $p = 0.643$). When we directly compared haploid and

diploid slopes within each environment, YPD ($t_{3.6}$ = 4.5, p=0.014), YPD+ethanol ($t_{5.93}$ = 4.28, p = 0.005) and YPD+NaCl ($t_{7.99}$ = 6.73, p = 0.0001) were found to differ significantly. The remaining four environments were not significantly different by this measure (YPD+KOH: t_6 = 1.24, p =0.261, YPD+HCl: $t_{7.3}$ = 1.85, p = 0.104, YPD+nystatin: $t_{5.28}$ = 1.62, p = 0.163, YPD+caffeine: $t_{7.18}$ = 1.37, p = 0.210).

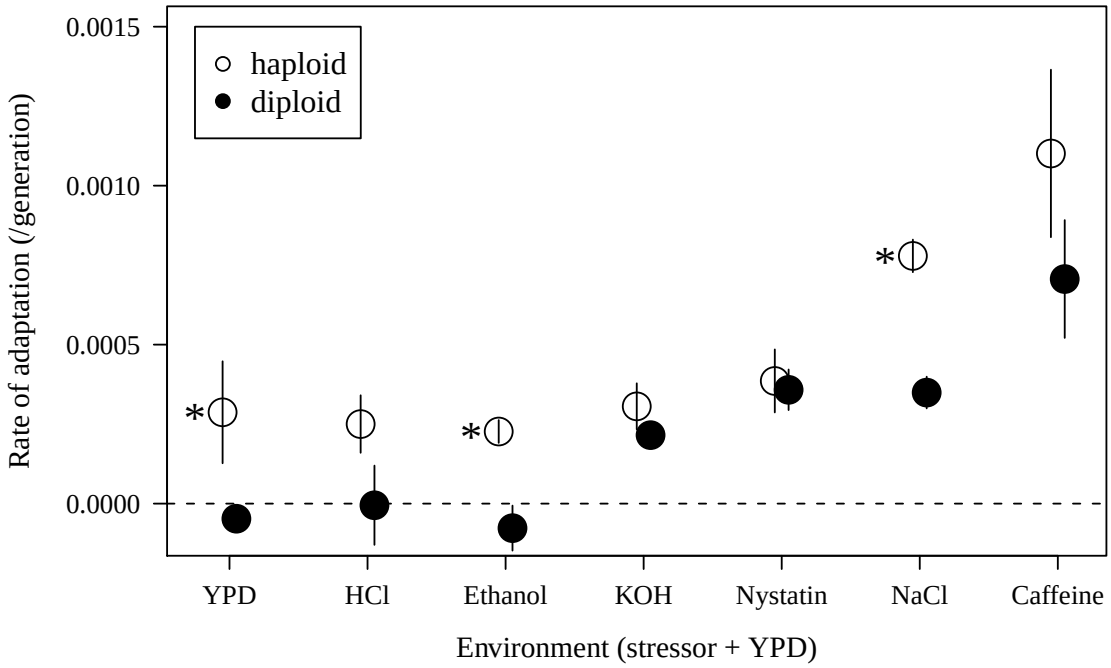


Figure 3.1: Haploids adapted faster than diploids in all environments. The rate of adaptation for haploid (open circles) and diploid (closed circles) was calculated as change in m (Malthusian parameter) over 140 generations. Each dot is the mean ±SE of five lines evolved independently. Stars (*) indicate a significant difference (p<0.05) between haploid and diploid lines (Welchs t-test)

.

We next measured the effective population size for all ancestral and evolved populations. As can be seen in Figure 3.2, the effective haploid population (open symbols) is generally greater than the effective diploid population (closed symbols) in all environments. We conducted a two-way ANOVA for each environment with time, ploidy and the interaction between them as predictors of population size. In all but two environments ploidy was the only significant factor (YPD: F_1 = 46.1, p<0.0001, YPD+HCl: F_1 = 11.1, p<0.0042, YPD+ethanol: F_1 = 62.8, p<0.0001, YPD+KOH: F_1 = 143.6, p<0.0001; see Table B.1 for full statistical results). In YPD+NaCl the evolved number of haploid cells decreased to that of the diploid lines and all three predictors were significant (ploidy: F_1=25.7, p=0.0001,

time: F_1=38.2, p<0.0001, ploidy*time: F_1=12.27, p=0.003). In contrast, none of the three predictors were significant in YPD+caffeine (ploidy: F_1=2.78, p=0.11, time: F_1=1.81, p=0.20, ploidy*time: F_1=2.02, p=0.17), driven by the tremendous variance of evolved haploid lines.

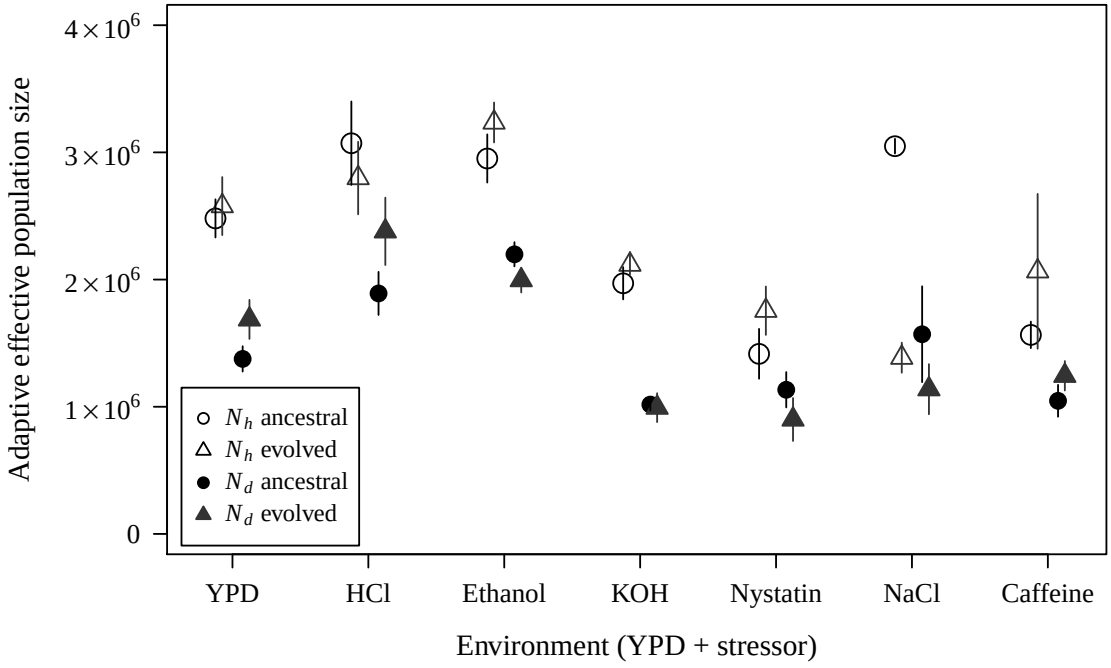


Figure 3.2: The effective population size of haploid (open symbols) and diploid (closed symbols) populations. In general, there was little difference between ancestral (circles) and evolved (triangles) population sizes.

.

For the remaining analyses, we split the environments into two groups, delineated by the diploid rate of adaptation. In the first group of environments (YPD+KOH, YPD+nystatin, YPD+NaCl, YPD+caffeine) diploids showed significant evidence of adaptation (one-sample t-test, μ=0; YPD+KOH: t_4= 6.40, p = 0.003, YPD+nystatin: t_4 = 6.47, p = 0.003, YPD+NaCl: t_4 = 7.84, p = 0.0014, YPD+caffeine: t_4 = -4.26, p = 0.013). In the second group (YPD, YPD+HCl, YPD+ethanol), we observed no evidence of diploid adaptation (one-sample t-test, μ = 0; YPD: t_3 = -1.07, p = 0.36, YPD+ethanol: t_4 = -1.22, p = 0.29, YPD+HCl: t_4= -0.05, p = 0.97).

In the first set of environments, we could infer the dominance of selected mutations in the diploid lines. Following from results in Otto and Whitton (2000) and using mathematical theory described in Appendix B, we used the ratio of the rate of haploid adaptation over the rate of diploid adaptation (Figure 3.1) and the measured effective population sizes (Figure 3.2) to estimate the dominance of beneficial mutations (equation B.3). As shown in

Figure 3.3, the dominance of selected mutations in all four environments was found to be consistent with partially dominant to overdominant mutations for fitness. We expect that only single mutations are present at high frequency in most lines (see Appendix B, Figure B.2), and thus this result describes the first selected mutations in each environment. These estimates were virtually unaffected by the mutation rate assumed (see Figure B.3; the rate used in the text was 10^{-7} beneficial mutations per genome per generation).

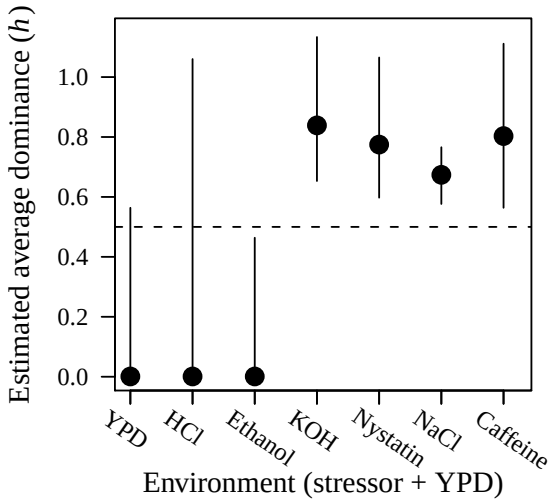


Figure 3.3: Mutations selected in diploid lines are predicted to be semidominant to overdominant in all environments where diploid lines adapted. Dominance estimates did not depend on whether we used ancestral (circles) or evolved (triangles) effective population sizes. Error bars indicate 95% confidence intervals obtained by parametric bootstrapping haploid and diploid rates of adaptation as well as haploid and diploid ancestral and evolved population sizes (see Appendix B for details).

.

That we observe diploid adaptation in one subset of environments but not in a second set suggests that something is different about the spectrum of mutations available in the different environments. In particular, we expect that either more mutations are available in the first set of environments (due to a higher mutation rate or a larger mutational neighbourhood) or that the mutation effect size of available mutations is greater. It is unlikely that the mutation supply rate was limiting because the population sizes were so large. Thus, we expect that either the effect size (s) and/or the dominance coefficient (h) of available mutations was lower in the second set of environments (YPD, YPD+HCl, YPD+ethanol). As discussed in Appendix B, we cannot exclude the possibility that dominance coefficients were uniformly high across all seven environments (Figure B.4).

3.5 Discussion

Consistent with theoretical expectations (Orr and Otto, 1994), we found that haploids adapted faster than diploids in seven different environments when evolved at large population sizes, significantly so in three environments. We expect that single mutations largely contributed to the improvement in fitness; as shown in Figure B.2, a fully dominant mutation that confers a 10% fitness advantage will only reach $\sim$ 50% frequency within 200 generations (see Appendix B). Assuming that single mutations contributed wholly to the observed adaptation, selection coefficients were found to be between 0 and 0.25 in the haploid lines (Table B.2). The mutations captured in our experiments are not necessarily representative of the distribution of all beneficial mutations. Due to clonal interference acting among multiple mutations appearing simultaneously in asexual populations (Rozen *et al.*, 2002), we expect mutations with higher s (haploid lines) or higher hs (diploid lines) to be most likely to spread to fixation. Previous microbial evolution experiments have found support for the presence of multiple mutations of moderate effect within populations (de Visser and Rozen, 2006; Desai *et al.*, 2007; Kao and Sherlock, 2008), which is also likely here. Keeping in mind that these are likely to be the best available mutations, our s estimates are consistent with other experiments performed in *S. cerevisiae* (Dickinson 2008: average s of beneficial mutations after 4800 generations of bottlenecks on YPD = 0.08, maximum of 0.12; Gresham *et al.* 2008: $s \approx$ 0.05-0.1 for beneficial mutations in carbon and phosphorus limitation; Desai *et al.* 2007: mean s = 0.02).

The results presented here describe a short term (<200 generation) evolution experiment. When we compare these to a previous study over a much longer scale ($\sim$1800 generations, Gerstein *et al.* 2006) we find a surprising disconnect between the rate of adaptation in the short term, and long-term shifts in ploidy. In this study we found that haploids adapted faster in both YPD and YPD+salt, yet in our previous study we saw that diploid mutants arose and took over all replicate lines within 1600 generations in YPD and 800 generations in YPD+salt. This contrast emphasizes the fact that simply accruing beneficial mutations at a faster rate does not protect haploid populations from invasion by diploid mutants during long-term evolution.

In four of the environments, the rate at which diploid lines adapted compared to haploids suggests that there are partially dominant beneficial mutations available to them. Although we did not observe significant diploid adaptation in the other three lines, we are unable to determine whether the dominance of available mutations differs significantly between environments (see Appendix B and Figures B.3 and B.4 for a theoretical discussion of why dominant mutations may still be available in these environments). Although we know from empirical results that the majority of deleterious mutations are partially recessive (Mukai *et al.* 1972, Ohnishi 1977, Mable and Otto 2001, Szafraniec *et al.* 2003,

and references within), we currently have few empirical estimates of the dominance of beneficial mutations. The most comprehensive study examined the dominance of beneficial mutations in pesticide and herbicide resistance genes. In a survey of more than 70 different studies, Bourguet and Raymond (1998) found that alleles that confer resistance via target site mutations varied from complete recessivity to complete dominance. The picture that emerges is that the average dominance of beneficial mutations depends on the particulars of the environment. Unfortunately, little is known about the dominance of beneficial mutations that arise under other selective pressures, in large part because it is difficult to isolate single beneficial mutations.

Yeast launched the genomics era of eukaryotes with the first published genome sequence in 1996 (Goffeau *et al.*, 1996), and yeast studies have continued to lead the charge in understanding the genomic basis of evolution (Dujon, 2010). Experiments such as these can be used not only to study population genetic questions but also to obtain testable predictions about the number and type of mutations that we may find as we move forward with broad-scale sequencing experiments. In particular, these results suggest that future sequencing studies should find mutations of larger effect size in haploid lines evolved in YPD+caffeine and YPD+NaCl compared to the other five environments. Similarly, we expect either mutations of larger effect size, or more dominant mutations in diploid lines evolved in YPD+caffeine, YPD+NaCl, YPD+nystatin and YPD+KOH. These experiments have demonstrated that haploids consistently evolve faster than diploids and suggest further experiments to confirm that the first steps of adaptation involve semi-dominant to dominant mutations in these environments.

Chapter 4

Parallel genetic changes and non-parallel gene-environment interactions underlie nystatin resistance in yeast

4.1 Summary

Beneficial mutations are required for adaption to novel environments, yet the range of mutational pathways that are available to a population has been poorly characterized, particularly in eukaryotes. We assessed the genetic changes of the first mutation acquired during adaptation to a novel environment (exposure to the fungicide, nystatin) in 35 haploid lines of *Saccharomyces cerevisiae*. Through whole genome resequencing we found that the genomic scope for adaptation was narrow; all adapted lines acquired a mutation in one of four late-acting genes in the ergosterol biosynthesis pathway, with very few other mutations found. Lines that acquired different ergosterol mutations in the same gene exhibited very similar tolerance to nystatin. All lines were found to have a cost relative to wild type in an unstressful environment; the level of this cost was also strongly correlated with the ergosterol gene bearing the mutation. Interestingly, we uncovered both positive and negative trade-offs between tolerance to nystatin and tolerance to other harsh environments for mutations in different genes, indicating that these beneficial mutations have effects that differ in sign among environmental challenges. These results demonstrate that although the genomic target was narrow, different adaptive mutations can lead populations down entirely different evolutionary pathways, with respect to their ability to tolerate (or succumb) to other environmental challenges.

4.2 Introduction

Populations adapt to stressful environments through the fixation of beneficial alleles. The number of advantageous mutations accessible to a population within one or a few mutational steps ("mutational neighbourhood", Burch and Chao 2000) remains poorly characterized especially in eukaryotes. This is an important factor, however, as the number of mutations in concert with their pleiotropic effects will directly influence the range of evolutionary pathways available to different populations. The first beneficial mutations to fix are of particular interest, as genetic and gene-environment (G×E) interactions may dictate the fixation of subsequent mutations. Knowledge of the number of available pathways may help us predict whether two populations subjected to similar selective pressures in allopatry might accumulate and fix different mutations. If this frequently occurs, reproductive isolation could evolve purely by chance fixation of different mutations (the mutation-order hypothesis, Schluter 2009). We thus sought to determine the mutational neighbourhood of adaptive mutations in one environment and to characterize the pleiotropic effects of these mutations to different environmental challenges to assess the extent of gene-environment interactions.

A fruitful approach to characterize the genotypic basis of adaptation has been experimental microbial studies, where multiple replicate populations are initiated with the same ancestral culture and evolved under the same conditions for up to thousands of generations (Conrad *et al.*, 2011). Targeted resequencing of specific genes in replicate populations evolved for hundreds or thousands of generations at large population size (where selection should overwhelm drift) has demonstrated that in many cases the same genes repeatedly acquire mutations (Barrick *et al.*, 2009; Ostrowski *et al.*, 2008; Woods *et al.*, 2006; Pelosi *et al.*, 2006; Cooper *et al.*, 2003). An examination of diverse clinical isolates of *Pseudomonas aeruginosa* has also repeatedly implicated the same genes during the acquisition of resistance to quinolone (Wong and Kassen, 2011). Parallel genotypic evolution is not restricted to the utilization of single genes, as parallel transposition mutations (Chou *et al.*, 2009) and large-scale aneuploid events (Selmecki *et al.*, 2009) have also been documented in replicate lineages evolved under the same conditions. Without resequencing the entire genome of an individual (through whole genome resequencing, WGS), however, it is unknown how many other (undetected) mutations contributed to adaptation in the prior mentioned studies. Furthermore, the magnitude of genetic parallelism is likely to be influenced by the selective environment (Anderson *et al.*, 2003; Gresham *et al.*, 2008), depending on both the size of the genomic target for beneficial mutations as well as the probability of establishment of different mutations.

A broader picture of the types of mutations acquired in long-term experimental evolution lines has been painted over the last few years by WGS. The results seem to suggest

that while the number of different genes available to adaptation and the types of mutations depends on both the species and the environment, some of the generalities from targeted re-sequencing studies seem to hold. In all cases there seem to be a small number of genes that are the target of independently evolved lines, and non-synonymous single nucleotide changes (SNPs) seem to be the most common type of mutation (Tenaillon *et al.*, 2012; Herring *et al.*, 2006; Kishimoto *et al.*, 2010; Minty *et al.*, 2011; Toprak *et al.*, 2011; Araya *et al.*, 2010). Considerable variation among experiments is found in the total number of different genes that are targeted, however, and variation is also present for the absolute frequencies of different classes of mutations (e.g., copy number variants, insertions, deletions, regulatory changes). However, such studies typically do not discern the order in which mutations appear, so the first step remains largely unknown, particularly for eukaryotes.

The order of mutational steps can have a tremendous impact on the fitness effect of subsequent mutations (Weinreich *et al.*, 2005), both in magnitude and sign, implying that the first adaptive step taken can alter the path of evolution. While WGS has allowed us to leap forward in our understanding of adaptation, fewer WGS studies have focused on the first step of adaptation, and none of these have yet characterized the first steps in a eukaryote. The data that exists suggest that the first mutations to be selected also tend to be clustered in relatively few genes, at least in viruses and prokaryotes. Rokyta *et al.* 2005 identified 10 unique non-synonymous single-step mutations through WGS within two different viral genes of ΦX174. Similar results were found through targeted re-sequencing in both *P. aeruginosa* (where 15 unique mutations were identified in *rpoB*, MacLean and Buckling 2009) and *P. fluorescens* (five non-synonymous SNPs were found in *gyrA* and four mutations were found within three efflux pump regulatory sites, Bataillon *et al.* 2011). As with all targeted resequencing studies, it remains unknown how many additional mutations were present in the lines acquired in *Pseudomonas*. It has also not yet been determined whether eukaryotes accumulate mutations in a similar (i.e., largely parallel) nature. A number of fundamental characteristics differ between prokaryotic and eukaryotic genomes (e.g., chromosome structure, the number of replication origins, the amount of non-coding DNA, and the degree of transcript processing, just to name a few, Poole *et al.* 2003). These or other factors could affect the nature of mutations acquired under stressful conditions in eukaryotic genomes.

To assess the mutational neighbourhood allowing adaptation to a novel stressful environment in a eukaryote we developed an assay to isolate multiple adapted lines of haploid *S. cerevisiae*. We exposed 240 replicate lines initiated with ∼100 000 progenitor cells to a level of stressor that inhibits growth of the ancestral strain. By immediately isolating cells that were able to grow in this environment, we limited the number of mutational hits in the genome and reduced the potential influence of clonal interference. Through WGS we pinpointed the genetic basis of adaptation for each lineage. We chose a polyene antibi-

otic, nystatin, as the stressful environment in which to acquire mutations. Nystatin binds to ergosterol (the primary sterol in the fungal membrane) to form porin channels that increase membrane permeability and allow cellular components (including potassium ions, sugar and metabolites) to leak out of the cell (Carrillo-Munoz *et al.*, 2006; Kanafani and Perfect, 2008). The resulting change in potassium concentration leads to an osmotic imbalance between the vacuole and cytoplasm and an enlarged vacuole (Bhiyan *et al.*, 1999). Transcriptional profiling has identified membrane transporters and the cell stress response as the major cellular components affected by exposure to nystatin (Hapala *et al.*, 2005). Previous work has identified resistant mutants in *Candida albicans* and *S. cerevisiae* that show defects in genes involved in the ergosterol biosynthesis pathway, with mutants exhibiting an altered sterol content in the cell membrane (Ghannoum and Rice, 1999; Bhiyan *et al.*, 1999; Kanafani and Perfect, 2008). Whether the ergosterol pathway would always be involved in resistance evolution, or whether beneficial mutations could be recruited from other membrane components or from altered ion pumps, remains unknown. The first goal of our research was thus to document the genetic basis of adaptation to nystatin in many replicate lineages and to measure the fitness benefit gained by each adapted lineage.

We then determined whether the mutations we identified exhibited differential responses to other stressful environments (G×E). Given that populations often face multiple environmental challenges simultaneously, the scope for adaptation would be greatly reduced if adaptive mutations in the presence of a single environmental change always exhibit reduced tolerance to other environmental challenges (strict trade-offs). The idea that strict trade-offs should exist during adaptation to novel environments is long-standing and multiple hypotheses have been put forward to explain the physiological basis of trade-offs (Pörtner *et al.*, 2006), yet trade-offs are not universally found (Hereford, 2009; Bennett and Lenski, 2007; Ostrowski *et al.*, 2005). Another possibility is that mutations vary in the subset of environments in which they are beneficial (i.e., there is "sign G×E", by analogy with sign epistasis, Weinreich *et al.* 2005). In this case, some mutations may be simultaneously beneficial to multiple types of change in the environment, allowing the organism to adapt more readily to complex environmental challenges. Furthermore, such gene-environment interactions imply that lineages carrying different first-step mutations would find themselves at different locations on the adaptive fitness surface after further changes in the environment. To explore the nature of gene-environment interactions among single adaptive mutations, we conducted a set of experiments to measure the fitness effects of different stressful conditions on mutations whose genetic basis is known, allowing us to compare mutations in different genes and at different sites within a gene. To explore a variety of environmental challenges, we varied levels of copper, ethanol, or salt, measuring growth of each line in each environment.

When exposed to nystatin, we found strong parallelism in the adaptive mutations that appeared within our *S. cerevisiae* lines at both the gene and pathway levels, with only a limited number of genes being involved in the first step of adaptation. Mutations in different genes had significantly different fitness effects across environments, with some lines showing increased tolerance and others decreased tolerance to other stressors (sign G×E). Our results thus provide support for the mutational-order hypothesis that adaptation to one environmental challenge may well drive isolated populations down different evolutionary pathways, with significant differences when faced with further environmental challenges.

4.3 Materials and Methods

4.3.1 *Strain background & mutation acquisition*

Mutations were acquired in haploids of genotype BY4741 (*MATa his3△1 leu2△0 met15△0 ura3△0*) derived from S288C. Stock of the ancestral culture and relevant gene deletion lines (see below) was ordered from Open Biosystems (Thermo Fisher Scientific), streaked on a YPD plate, and a single colony was isolated and frozen. Thus, a single colony from BY4741 was chosen as the progenitor culture for all experiments, and we refer to this as the ancestral strain. All experiments were carried out in 1mL culture in 96 deep-well plates (2mL polypropylene plates with a conical bottom), shaking at 200rpm at 30°C. We acquired mutations in two screens, separated by two weeks. To initiate each screen, the ancestral strain was streaked onto a YPD plate from frozen and grown for 48 hours. A single random colony was then picked and grown for 24 hours in 10mL YPD at 30°C, shaking at 200rpm.

We initiated mutation acquisition in the first screen by transferring 10 μL of the ancestral overnight culture into 1mL of YPD+4μM nystatin in the 60 inner wells of a 96 deep-well plate. The second screen was identical except we initiated 180 replicates into the inner wells of three 96 deep-well plates. The level of nystatin was determined in preliminary experiments as the level that showed only sporadic growth of the ancestral culture within 2-7 days of incubation, suggesting that growth required a mutational event (preliminary results not shown). Growth in nystatin was checked and recorded daily by visual examination of the bottom of the 96 well plates. A small amount of growth would typically be observed one day, with full growth on the second or third day (where full growth is approximately equivalent to the amount of turbidity and precipitate of the ancestral culture in YPD overnight). Occasionally full growth took up to 4 days. Each well that showed growth (even slow growth) was marked as a "putative mutation" line.

On the first day that full growth was recorded, the well was thoroughly mixed by pipetting and culture was streaked onto a YPD plate. Putative beneficial mutation lines were obtained in this way from 73 out of the 240 inoculated wells, with no growth observed

in the remaining wells over seven days. After 48 hours of growth on the plate, we visually assessed each line for petite mutations (mutations that affect mitochondrial function and prevent respiration, these colonies present as much smaller than normal on a YPD plate). About half of the putative mutation lines showed evidence of petite mutations. All suspected petites were confirmed by lack of growth on a YPG plate, a medium that requires respiration for growth, and these lines were discarded to focus on nuclear mutations. For each of the remaining 35 lines, eight colonies were haphazardly picked off the YPD plate, placed back into eight wells containing 1mL of YPD+4μM nystatin and assayed for growth. Often all eight colonies picked would exhibit similar growth patterns, but sometimes colonies varied with respect to the number of days to full growth or even whether any growth was observed within 48 hours. To avoid analyzing non-mutant cells that might have been segregating, we randomly picked a single well that showed any growth in nystatin for each mutation line. The 1mL culture from this well was mixed with 1mL 30% glycerol and frozen; this freezer culture constitutes the material for all future experiments. These 35 lines were labelled BMN1–35 (BMN: "Beneficial Mutation Nystatin") as described in the Table 4.1 legend.

4.3.2 *Sequencing*

Freezer culture from each BMN line was streaked onto YPD plates and grown for 48 hours. A single colony was then haphazardly picked for each line and grown for 24 hours in YPD. DNA was extracted (Sambrook and Russell, 2001) and sequenced in 100bp single-end fragments using Illumina's HighSeq 2000. Library preps followed standard Illumina protocols (2011 Illumina, Inc., all rights reserved), with twelve uniquely barcoded strains run together on a lane. The resulting genomic sequence data were processed using Illumina's CASAVA-1.8.0. Specifically, configureBclToFastq.pl was used to convert to fastq and separate the sequences by barcode (allowing one mismatched basepair). configureAlignment.pl (based on the alignment program ELAND) was then used to align each sequence to the yeast reference genome (scergenome.fasta downloaded from the Saccharomyces Genome Database http://*downloads.yeastgenome.org/genome_release/r*64/).

SNPs and small insertions and deletions (indels) were then called using configureBuild.pl. Average coverage per mapped site across the strains (excluding mitochondrial genes) was 44.0 (with a minimum average coverage per site of 4.7 for BMN13). Data on average coverage per chromosome is illustrated in Figure C.1, which indicates that one line (BMN27) had an additional copy of chromosome 2 (denoted as a star in Figure C.1). Custom UNIX and perl scripts were then used to parse the output files. Illumina data from an independent project using the same ancestral line were used to identify mutations that were common to the ancestor, and all such mutations were ignored. Given that our initial lines

were haploid, variants identified by configureBuild.pl as heterozygous were also discarded as likely alignment or sequencing errors. Similarly, variants involving repeat elements were discarded. All remaining variants were checked in the alignments using tview in samtools-0.1.7a (Li *et al.*, 2009); variants that were not supported by multiple fragments starting from different positions were also discarded (typically near deletions or gaps in the alignment). Finally, the same procedure was repeated, but using the bwa software package to perform the alignment (Li *et al.*, 2009) along with samtools-0.1.7a to identify SNPS (Li *et al.*, 2009), using the -bq 1 option to limit data to reliable alignments. All SNPs (Table 4.1 & Table C.1) were identified using both methods.

Sequence alignments were manually checked using tview in samtools-0.1.7a for the four genes harbouring beneficial mutations (*ERG3*, *ERG5*, *ERG6*, *ERG7*) to look for larger rearrangements or other changes not identified by the above procedure. Two additional large-scale mutations were identified from gaps in the alignments. To determine the nature of these rearrangements, the fastq files containing the unaligned short-read sequences were directly searched for sequences on either side of the alignment gap, confirming a 60 bp deletion in *ERG5* within line BMN35 and a 29 bp duplication in *ERG3* within line BMN28 (Table 4.1). To confirm the accuracy of this method, we Sanger sequenced the appropriate gene from BMN lines representing 10 of the 20 unique mutations found (Table 4.1). In all cases the Sanger sequence data matched our analysis of the genomic sequence data. Importantly, our assays for mutations will miss larger rearrangements not involving these four genes, as well as any mutations occurring within repetitive sequences, which were ignored.

4.3.3 *Sterol assay*

We compared the sterol profile of the ancestral strain (BY4741) and BMN lines using a spectrophotometry-based assay. When more than one BMN line shared the same ergosterol mutation we randomly chose one line to represent that group. Sterols were extracted using the alcoholic potassium hydroxide method as previously described (Arthington-Skaggs *et al.*, 1999). BMN culture streaked to single colony on a YPD plate was inoculated into 50mL of YPD and grown at 30° for 48 hours at which point the optical density (OD) was measured to record cell concentration. Cells were then harvested by centrifugation at 2700 rpm for 5 minutes and washed twice with sterile distilled water. 3mL of 25% alcoholic KOH was added to each pellet and vortexed for 1 minute. The sample was then incubated in an 80°C water bath for 1 hour then cooled to room temperature. To extract the sterols, 1mL of sterile distilled water and 3mL of heptane were added and vortexed for 3 minutes. A 200μL aliquot of the heptane layer was added to 800μL of 95% ethanol, and the absorbance was immediately read every 3nm between 200 – 300nm with a Thermo BioMate 3 spectrophotometer.

4.3.4 *Nystatin tolerance*

To determine the breadth of nystatin tolerance conferred by each mutation, a growth assay was performed to measure the half maximal inhibitory concentration (IC_{50}) of nystatin. Freezer stock from each BMN line was inoculated into 96 deep-well plates containing 1mL of YPD and grown for 48 hours. To standardize the starting density of cells in each culture, the OD of 200μL from each culture was measured using the BioTek microplate reader (BIoTeck Instruments, Inc. Winooski, VT) and diluted to the sample with the lowest OD (usually between 0.7-0.9 OD630nm). 200μL of the standardized culture was then added to 400μl of YPD to obtain the final volume necessary for the assay inoculations. For each BMN line, 12μL of the dilute culture was then inoculated into a well containing 1mL YPD plus one of ten levels of nystatin (specifically: 0μM, 2μM, 4μM, 8μM, 12μM, 16μM, 46μM, 96μM, 116μM, 200μM), each replicated in four different wells. Plates were sampled at 72 hours to measure OD. Wells were manually mixed and 150μL aliquots were taken from each well and read on the BioTek reader.

A maximum likelihood model was fit to the data to determine IC_{50}. The logistic function

$$y = \frac{y_{max} exp(a(x - IC_{50}))}{1 + exp(a(x - IC_{50}))} + N(0, \sigma) \tag{4.1}$$

was used, where x represents the tested concentration of nystatin, y represents the observed OD following 72 hours of growth, and N(0, σ) represents a normal deviate with mean zero and standard deviation, σ. The fitted parameters were y_{max} (the maximal OD under full growth), IC_{50} (the nystatin concentration at which OD is half maximal), and a (the slope of the logistic curve at $x = IC_{50}$ divided by $y_{max}/4$), and σ. The maximum likelihood point was found in R using the subplex method of optim, as implemented in the find.mle routine of the diversitree package (FitzJohn *et al.*, 2009). The find.mle routine allows lower and upper limits to the parameters to be specified in the search routine (we used lower: y_{max}=0.8, IC_{50}=0.0000001, a=-50, σ=0; and upper: y_{max}=1.2, IC_{50}=6, a=0, σ=10; upper y_{max} and IC_{50} were based on observations). Prior to fitting the data using this likelihood procedure, all nystatin concentrations (and IC_{50}) were ln-transformed (so that percentage changes, not absolute differences, in nystatin matter), but all reported values are on the original scale.

To determine whether the IC_{50} of a mutant line was significantly different from the ancestral strain, a likelihood model was fit to the data from the mutant line and the ancestral line, allowing each of these two lines to have its own values of y_{max}, IC_{50}, a, and σ. This "full" model was then compared to a constrained model where $IC_{50,mutant} = IC_{50,ancestral}$ using a likelihood ratio test. If the drop in log-likelihood between the full and constrained model was greater than $\chi^2_{1,0.05}/2 = 1.92$, we rejected the hypothesis that IC_{50} was the same for the two lines.

4.3.5 *Fitness proxies in a permissive environment and the evolutionary environment*

We assessed growth in both the evolutionary environment (YPD+4μM nystatin) and an unstressful environment (standard lab YPD) using two fitness proxies. We measured optical density at 48 hours to capture the ability to turn resources into biomass and maximal growth rate to measure how quickly yeast cells are taking in nutrients and growing during the exponential phase of growth. Both fitness proxies were determined using previously described methods (Chapter Two, Gerstein and Otto 2011) that utilize the Bioscreen C Microbiological Workstation (Thermo Labsystems), which measures optical density (OD) in 100-well honeycomb plates. In brief, plates were streaked from frozen stock onto YPD plates for all lines and the ancestral culture and allowed to grow for 72 hours. Inoculations containing one colony (haphazardly chosen) for each BMN line and five separate inoculations of the ancestral strain (each from a different single colony) were then allowed to grow for 48 hours in 10mL YPD. 100μL was transferred into 10mL of fresh YPD, mixed thoroughly, and four 150μL aliquots for each line were placed into non-adjacent bioscreen wells. The bioscreen plates were grown at 30° for 48 hours, with constant shaking; OD readings were automatically taken every 30 minutes. We determine the maximal growth rate for each well as the spline with the highest slope, from a loess fit through log-transformed optical density data using an analysis program written by Richard Fitzjohn in R (R Development Core Team, 2011). In addition, OD at 48 hours was used as a second fitness measure ($OD48$). As can be seen from the raw growth curves (Figure C.2) the lines have stopped growing by 48 hours in YPD, and this measure thus reflects efficiency (i.e., ability to turn nutrients into cellular material). In nystatin, by contrast, some lines may still be growing, and this assay thus represents a combined measure of growth rate and efficiency.

4.3.6 *Assessing gene-environment interactions*

The ecological tolerance (measured as IC_{50}) was determined for each line in copper ($CuSo_4$), ethanol, and salt (NaCl). The tolerance assays in these environments were conducted as previously described for nystatin; we measured growth after 72 hours in eight levels of copper (0mM, 1mM, 2mM, 4mM, 6mM, 8mM, 10mM, 12mM), seven levels of ethanol (0%, 2%, 4%, 6%, 8%, 12%, 14%), and eight levels of salt (0.2M, 0.4M, 0.6M, 0.8M, 1.2M, 1.4M, 1.6M, 2M). These levels were chosen based on preliminary data that indicated the approximate position of IC_{50}. Significance was determined as with tolerance to nystatin, using a maximum likelihood test that compared the model fit with two IC_{50} parameters (one for the mutation line and ancestor) to a model with only one IC_{50} value.

4.4 Results

We acquired 63 haploid lines of *S. cerevisiae* that were resistant to nystatin in two independent screens. We eliminated all lines that could not respire, which left us with 35 lines that we term BMN lines ("beneficial mutation nystatin"). Through whole genome sequencing (WGS) with the Illumina platform (Illumina, San Diego, CA), we found that each line carried a single mutation in one of four late-acting genes in the ergosterol biosynthesis pathway (Figure 4.1). Within these lines, we found one line each with mutations in *ERG7* ("BMN-*erg7*") and *ERG5* ("BMN-*erg5*"), seven unique mutations in *ERG6* within 19 lines (collectively referred to as "BMN-*erg6* lines"), and eleven unique mutations within 14 lines in *ERG3* ("BMN-*erg3* lines"). We recovered multiple classes of mutations including nonsynonymous SNPs, premature stop codons, small indels (<3 basepairs), one 60bp deletion, and one 17bp duplication (Figure 4.1); the precise nucleotide and amino acid changes as well the numbering scheme for BMN lines are presented in Table 4.1.

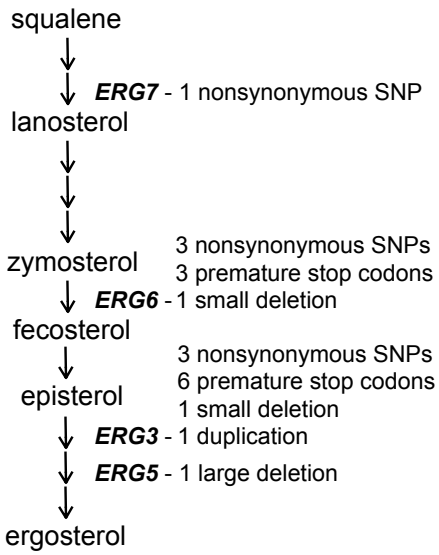


Figure 4.1: Twenty unique mutations were found in four late-acting genes in the ergosterol biosysnthesis pathway Each arrow in this figure represents one gene in the pathway that converts squalene to ergosterol. See Table 4.1 for detailed information on the genetic nature of each mutation.

For five mutations, the same sequence change was observed in multiple lines (Table 4.1). There are three potential explanations for this finding. The most likely is that mutations were initially segregating in the source population before the stressor was applied (see Appendix C). A second explanation, that contamination occurred across wells in the 96 well plates, is possible, but fails to explain the similar timing of appearance of identical hits

Table 4.1: The genetic basis of BMN line ergosterol mutations.

Line (BMN)	Gene	Genome Position (Chr.Bp)	Position in Gene (in nucleotides)	Mutation	Amino Acid Change
1	*ERG7*	XIII.241194	2096	C>G[a]	Phe699Leu
2-4	*ERG6*	XIII.252861	131	G>A[a]	Gln44Stop
5	*ERG6*	XIII.252772	220	G>T[a]	Tyr74Stop
6	*ERG6*	XIII.252723	269	G>T	His90Asn
7-10	*ERG6*	XIII.252612	380	C>G[a]	Gly127Arg
11-15	*ERG6*	XIII.252596	395	CC/–[a]	
16	*ERG6*	XIII.252349	642	C>G	Leu214Phe
17-20	*ERG6*	XIII.252322	669	G>C	Tyr223Stop
21	*ERG3*	XII.254048	187	A>T[a]	Arg63Stop
22	*ERG3*	XII.254088	227	C>A	Ser76Stop
23	*ERG3*	XII.254144	284	C>A[a]	Ser95Stop
24-27	*ERG3*	XII.254475	615	G>A	Trp205Stop
28	*ERG3*	XII.254500	640	29bp duplication[b]	
29	*ERG3*	XII.254516	656	G>A	Trp219Stop
30	*ERG3*	XII.254563	703	G>A	Gly235Arg
31	*ERG3*	XII.254757	897	C>A	Tyr299Stop
32	*ERG3*	XII.254758	898	G>Ca	Gly300Arg
33	*ERG3*	XII.254781	920	A>C	Asp307Ala
34	*ERG3*	XII.254840	980	A/-	
35	*ERG5*	XIII.301120	252	60bp deletion[c]	

Each BMN line carried a single mutation in one of four genes at the end of the ergosterol pathway. We numbered each line sequentially based on the location of each ergosterol mutation. Lines with mutations in genes furthest in the pathway from producing ergosterol (the end product) have lower numbers; within a gene, mutations nearer the start codon were given lower numbers.

[a] Mutations were confirmed with Sanger sequencing (when multiple lines shared a mutation we confirmed the mutation in only a single line; BMN3, BMN9, and BMN13)

[b] Confirmed with Sanger sequencing – bp640-669 are duplicated and inserted after bp669

[c] Confirmed with Sanger sequencing – deletion of 60 bases between bp252–bp312

(Table C.6). The independent appearance of the same sequence changes is also possible, but in no case did we observe the same mutation in the two screens (Table C.6). While we treat each line as independent for statistical purposes, we note that combining lines with the same ergosterol mutation leads to the same conclusions (Appendix C.2).

We did not expect to see many mutations other than those conferring a fitness benefit given the relatively small initial population size ($\sim 10^5$ cells), small genome-wide mutation rate (Lynch *et al.*, 2008), and short time frame of the experiment (we stopped once growth could be observed, thus minimizing the number of generations, Appendix C.3). Nevertheless, we identified a small number of additional point mutations in genes that are not part of the ergosterol pathway (15 in total across the 35 lines; Table C.7). The majority of these mutations were unique to a single BMN line, but two mutations were found in multiple lines. A nonsynonymous change from glutamic acid to lysine in *FCY2* was found in four lines, and a synonymous mutation was found in *GDA1* in five lines; in both cases, these two sets of lines also shared a primary ergosterol mutation (Table 4.1), strongly suggesting that these two sets may be derived from the same mutations that arose in the precursor population (as previously discussed). The genome size of all lines was measured using flow cytometry, and no deviations from haploidy were found. Examining the depth of coverage from Illumina data (see Methods) also uncovered one case of chromosomal aneuploidy (BMN27 had a duplicated chromosome II, Figure C.1). We did not find strong evidence that any of the non-ergosterol mutations influence fitness in the environments measured (see Appendix C), and thus we focus our discussion on the ergosterol mutations.

We first measured the sterol profile of all lines. This assay takes advantage of the characteristic four-peak curve produced by ergosterol and the late sterol intermediate 24(28)dehydroergosterol (DHE) that are present in wild-type cells (Arthington-Skaggs *et al.*, 1999). All lines that carried mutations in the same ergosterol gene showed nearly identical sterol profiles (measured using a spectophotometry based assay, Figure 4.2). Interestingly, only BMN-*erg5* (the line with a mutation in the gene closest to the end of the pathway) had a sterol profile similar to the ancestral strain. The sterol profiles for BMN-*erg6* lines and BMN-*erg3* lines have a similar shape to previously published results of *erg3*Δ and *erg6*Δ obtained using the same protocol (Jensen-Pergakes *et al.*, 1998). Surprisingly, however, our own measures of *erg6*Δ and *erg3*Δ (and *erg5*Δ) recovered the ancestral sterol phenotype (not shown).

All mutation lines had a significantly higher tolerance to nystatin than the ancestral strain, and many lines could tolerate nystatin at much higher levels than the 4μM concentration used to isolate beneficial mutations (Figure 4.3). We measured the breadth of tolerance as IC_{50}, i.e., the inhibitory concentration of the drug that reduced growth by 50%. The significance of changes in IC_{50} relative to the ancestor was determined by maximum likelihood (described in the Methods, results in Table C.8). Replicate lines that carried mu-

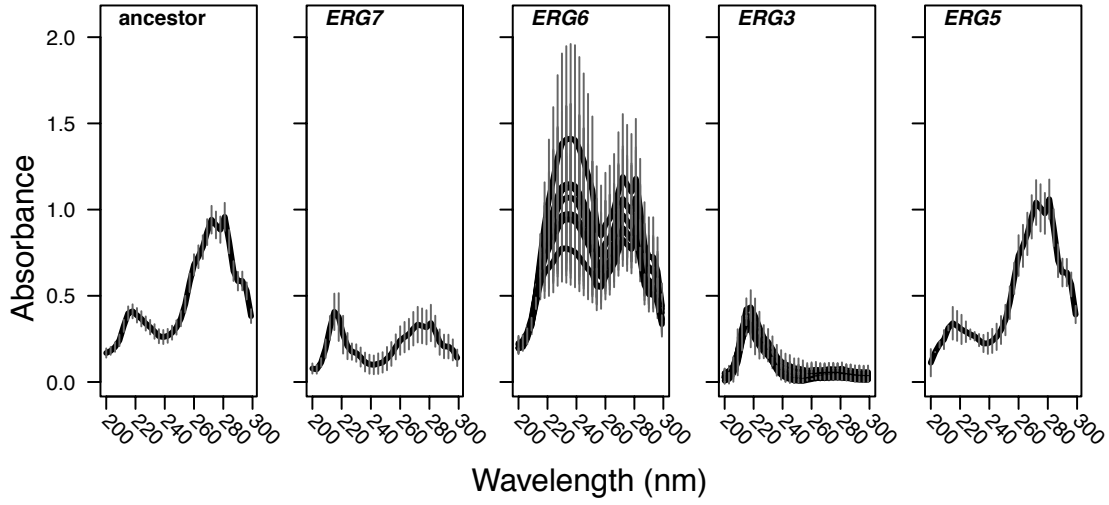


Figure 4.2: The sterol profile of all BMN lines is different from the ancestral profile except for the line with a mutation in *ERG5*. Sterol profile of each line was measured using a spectometry based assay. For BMN lines that carry the identical ergosterol mutation, a single line was randomly chosen to represent the group (BMN-*erg6*: BMN3, 6, 9, 12, 16-18; BMN-*erg3*: BMN25, 27, 29, 30-33). Error bars depict the standard error of replicates measured on three different days.

tations in the same ergosterol gene showed similar IC_{50} values (Figure 4.3). A two-way ANOVA found that IC_{50} in nystatin has a very strong association with the ergosterol gene bearing a mutation ($F_3 = 252.4$, $p < 0.0001$) but was not affected by either the class of mutation within a gene (i.e., non-synonymous SNP, premature stop codon, or indel) or their interaction (mutation type: $F_2 = 0.92$, $p = 0.41$; interaction: $F_2 = 0.66$, $p = 0.53$). We then compared the tolerance of our lines to *S. cerevisiae* strains that carry gene knockouts for *ERG6*, *ERG3* and *ERG5* (*erg7*Δ is inviable and could not be tested). Although all gene knockout lines did show increased nystatin tolerance compared to the ancestor (Figure 4.3), we found that while BMN-*erg5* and *erg5*Δ had similar IC_{50} values (Figure 4.3), BMN-*erg6* lines had a significantly higher nystatin tolerance than *erg6*Δ, and all but two BMN-*erg3* lines had a significantly lower nystatin tolerance than *erg3*Δ.

We also measured two fitness-related proxies for all lines in both the evolutionary environment (YPD+4μM nystatin) and an unstressful environment (standard lab YPD). When grown in nystatin all BMN lines reached a higher optical density by 48 hours ($OD48$, Figure 4.4A) and had a higher maximal growth rate (Figure 4.4C) than the ancestral strain (significance determined by a t-test compared to five ancestral colonies, Figure 4.4 & Tables C.9-C.12). Conversely, the ancestor performed better than all BMN lines in YPD for both fitness proxies (Figure 4.4B & D). Growth rate and $OD48$ were significantly correlated to each other when BMN lines were grown in nystatin, consistent with the idea that both

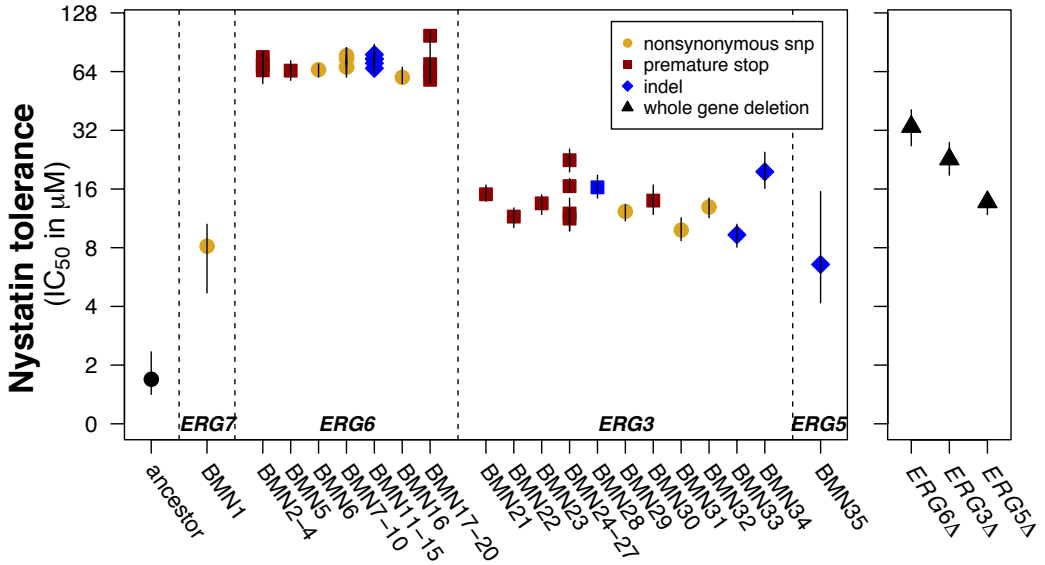


Figure 4.3: BMN lines have significantly increased tolerance to nystatin relative to the ancestor. Error bars represent 95% confidence intervals from likelihood profile plots. Lines grouped on the x-axis carried identical ergosterol mutations at the sequence level (Table 4.1). Mutation line are arranged on the x-axis in the same way they were numbered, i.e., based on which gene carries a mutation (mutations in genes further from producing ergosterol are numbered lower and plotted to the left), and position in the gene (mutations closer to the start codon are numbered lower and plotted to the left of mutations closer to the stop codon).

assays measure an aspect of growth rate (cor = 0.87, $t_{333} = 10.3, p < 0.0001$), and both were significantly correlated with nystatin IC_{50} ($OD48$: cor = 0.74, $t_{33} = 6.4, p < 0.0001$; growth rate: cor = 0.76, $t_{33} = 6.8, p < 0.0001$). When lines were grown in YPD, however, the two growth measures were not significantly correlated with each other, consistent $OD48$ measuring the efficiency of converting resources into cellular material (cor = 0.26, $t_{33} = 1.5, p = 0.13$). Growth rate in YPD was significantly correlated with IC_{50} in nystatin (cor = 0.56, $t_{33} = 2.8, p = 0.01$), while $OD48$ was not (cor = 0.04, $t_{33} = 0.18, p = 0.20$). Interestingly, the correlation between growth rate in YPD and IC_{50} in nystatin was positive; that is, mutations with the broadest tolerance to nystatin were among the best to grow in YPD. The ergosterol gene that bore a mutation was significantly associated with both fitness proxies in both environments (growth rate in nystatin: $F_3 = 26.4$, $p < 0.0001$; $OD48$ in nystatin: $F_3 = 15.5$, $p < 0.0001$; growth rate in YPD: $F_3 = 25.8$, $p < 0.0001$; $OD48$ in YPD: $F_3 = 4.9$, $p = 0.007$). The type of mutation (i.e., non-synonymous SNP, premature stop codon, or indel) was also found to have a significant effect on growth rate in nystatin ($F_2 = 5.0$, $p = 0.014$), though we note that of all significant statistical results this is the only one that does not remain significant when we combine multiple lines with the same ergosterol mutation (see Appendix C). In all other comparisons we found no significant association with the type of mutation ($OD48$ in nystatin: $F_2 = 0.86$, $p = 0.44$; growth rate in YPD: $F_2 = 0.27$, $p = 0.77$; $OD48$ in YPD: $F_2 = 0.09$, $p = 0.91$) nor was the interaction between ergosterol gene and type of mutation significant (growth rate in nystatin: $F_2 = 0.16$, $p = 0.85$; $OD48$ in nystatin: $F_2 = 0.68$, $p = 0.51$; growth rate in YPD: $F_2 = 0.33$, $p = 0.72$; $OD48$ in YPD: $F_2 = 0.73$, $p = 0.50$).

We observed substantial differences among the nystatin resistance lines in their tolerance to other stressful environments (breadth of tolerance measured as IC_{50} in all environments, Figure 4.5). We found significant negative correlations between tolerance to nystatin and tolerance to both copper and ethanol (copper: cor = -0.80, $t_{34} = -7.6$, $p < 0.00001$; ethanol: cor = -0.63, $t_{34} = -4.7$, $p < 0.00001$), and no correlation between nystatin and salt tolerance (cor = -0.07, $t_{34} = -0.39$, $p = 0.70$). The tolerance breadth exhibited by lines with different mutations in the same ergosterol gene were fairly consistent, with only a few exceptions. Importantly, although reduced tolerance to all other environments tested was observed for some genes bearing nystatin resistance mutations (especially *ERG6* mutations), mutations in other ergosterol genes had no effect or even a positive effect on growth in the face of other environmental challenges (e.g., positive fitness effects were observed for BMN-*erg7*, BMN-*erg3*, and BMN-*erg5* lines in copper). That is, mutations in different ergosterol genes exhibited significant sign G×E when comparing growth in nystatin and copper. Although the majority of lines differed in fitness from the ancestor in most environments (Figure 4.5, Table C.13), the pattern was also heavily dependent on both environment and gene.

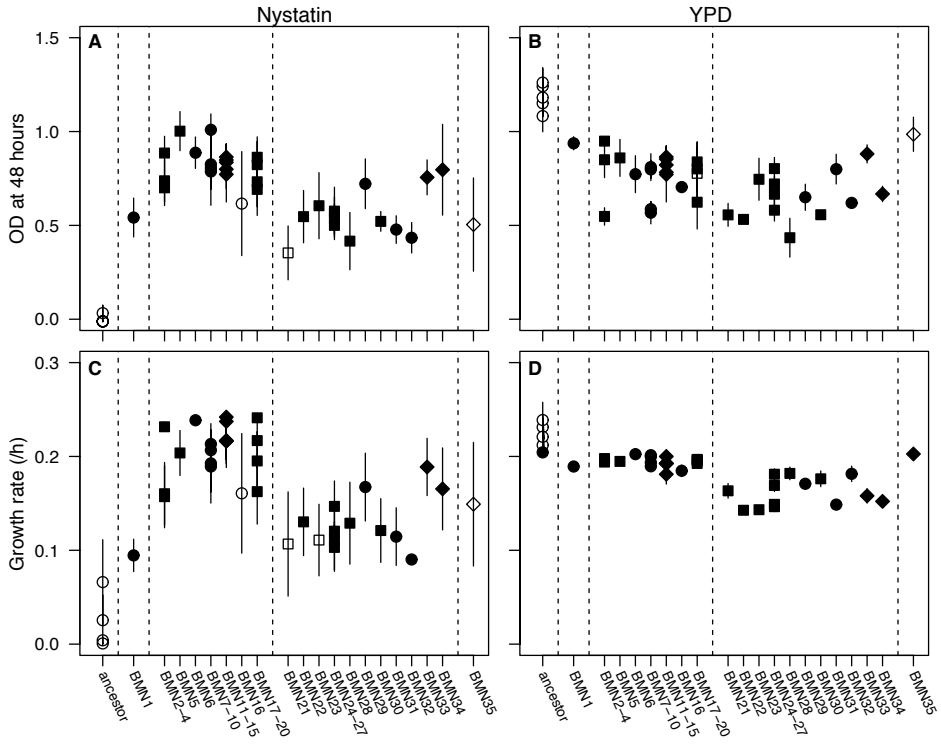


Figure 4.4: BMN lines show increased growth in the evolutionary environment and reduced growth in an unstressful environment relative to the ancestor. Two different fitness assays, OD at 48 hours (A - in nystatin; B - in YPD, an unstressful environment) and maximum growth rate (C - in nystatin, D - in YPD) show that BMN lines have increased growth in YPD+4μM nystatin and reduced growth in YPD relative to the ancestor. Lines that are significantly different than the ancestral colonies are plotted with closed symbols (t-test results presented in Tables S4-S7). Error bars depict the standard error of four bioscreen well replicates for each colony.

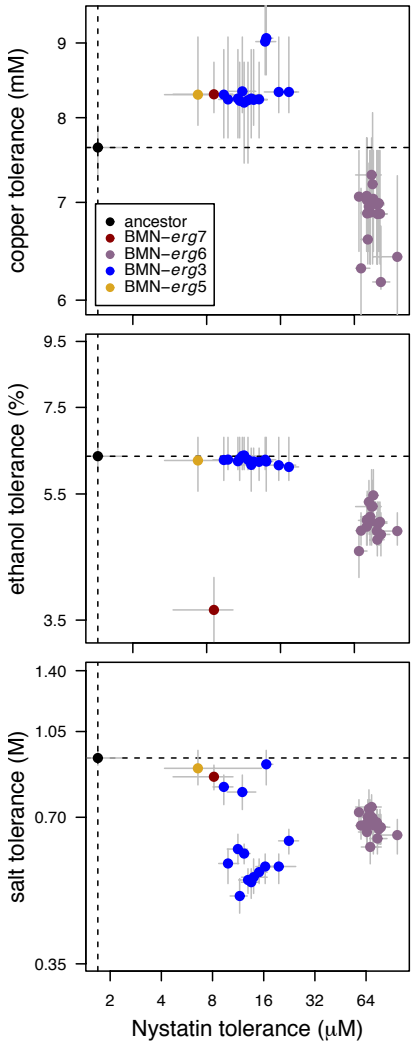


Figure 4.5: Tolerance to nystatin does not predict tolerance to other stressful environments. The tolerance (measured as IC_{50} in all environments) was measured in copper ($CuSO_4$), ethanol and salt (NaCl). Dashed lines indicate the ancestral tolerance to each environment. Lines that appeared to have very similar tolerance to one environment did not necessarily have a similar tolerance to a second, indicating significant G×E interactions.

4.5 Discussion

Genes that act late in the ergosterol biosysnthesis pathway were found to be the primary (and possibly exclusive) target for the first step of adaptation by *S. cerevisiae* to low levels of nystatin. The mutational neighbourhood was reasonably large, as we uncovered twenty unique mutations within four genes exhibiting increased tolerance to nystatin. At the gene level, however, the genomic scope for beneficial mutations was quite narrow in this environment, as all but two lines carried mutations within *ERG6* and *ERG3*. Lines with different mutations in the same gene tended to exhibit similar tolerance phenotypes in all environments tested, including altered levels of the original stressor (nystatin), an unstressful environment (YPD), and three different stressful environments (copper, ethanol, and salt). We found that mutations in different ergosterol genes had non-parallel fitness effects in the face of other environmental challenges, indicating the unpredictable nature of gene-environment interactions. Although some lines showed a decreased fitness in all other stressful environments tested (i.e., BMN-*erg6* lines), other lines showed a mixture of fitness costs and benefits in other environments, with some lines having high tolerance in all other environments tested (i.e., BMN-*erg5*).

Parallel evolution is more likely to occur via loss of function mutations than gain of function (Christin *et al.*, 2010), and the different mutations that we observed in *ERG6*, *ERG3* and *ERG5* may well have caused loss of function alleles. Consistent with this hypothesis, knockout lines for these three genes (*erg6*Δ, *erg3*Δ, and *erg5*Δ) are viable and have been shown to increase fitness in low levels of nystatin in a screen of all deletion collection lines (Hillenmeyer *et al.*, 2008). Our own nystatin tolerance assay of these null mutations in nystatin showed subtle but significant differences between BMN-*erg6* and BMN-*erg3* lines and the appropriate knockout lines (Figure 4.3). These results suggest that our mutations are similar, but not identical, to the null mutations, and may indicate that the enzymes these genes encode may retain some activity. Different amino acid changes in *ERG6* have previously been shown to have different kinetic properties (Nes *et al.*, 2004), and so we must be cautious in concluding that these mutations all represent complete loss of function. By contrast to *ERG3* and *ERG6*, we only identified one mutation in *ERG5*. This lack of parallel mutations in *ERG5* is somewhat surprising, given that this is a longer gene (1616bp) than either *ERG6* (1151bp) or *ERG3* (1097bp). One possible explanation is that as we only characterized non-petite mutations to avoid mitochondrial mutations, we may have missed other mutations in *ERG5*, as *erg5*Δ is respiratory deficient (Merz and Westermann, 2009).

In contrast to all other lines, BMN1 (with a mutation in *ERG7*) cannot carry a loss of function mutation, as *erg7*Δ is inviable. Consistent with this claim, the only mutation that arose in *ERG7* was a nonsynonymous change very close to the end of the gene (BMN1). It is thus plausible that this particular change was a gain of function mutation. Furthermore,

our sterol profile of BMN1 is unique among the mutations we acquired. Similarly, BMN1 does not seem to share a sterol profile with any of the mutants identified in earlier studies on nystatin resistance, though sterol profiles in these early mutants match the profiles exhibited by our lines with mutations in *ERG6*, *ERG3* and *ERG5* (Woods, 1971; Bard, 1972; Grunwald-raij and Margalith, 1990). The unique phenotype generated by the mutation in *ERG7* deserves future investigation. It may be that other gain of function mutations could have been beneficial in nystatin but were not sampled due to rarity or a bias towards large effect loss-of-function mutations in our screens. Interestingly, *S. cerevisiae* knockout lines for all other genes that act late in the ergosterol pathway (except *hmg1*Δ and *hmg2*Δ, which are isozymes, so that deleting either one alone is not expected to have a strong effect on growth), are are inviable (*erg10*Δ, *erg13*Δ, *erg12*Δ, *erg8*Δ, *mvd1*Δ, *erg20*Δ, *erg9*Δ, *erg1*Δ, *erg11*Δ, *erg25*Δ, Giaver *et al.* 2002), unable to grow aerobically under our growth conditions (*erg24*Δ, Lees *et al.* 1995), ergosterol auxotrophs (*erg2*Δ, Parks and Casey 1995), or have reduced fitness in nystatin (*erg4*Δ, Hillenmeyer *et al.* 2008), helping to explain the narrow gene target of adaptation to nystatin that we have observed.

The distribution of fitness effects of beneficial mutations is an important factor that dictates how populations might adapt to a novel stressor. In his seminal work on adaptive mutations, Gillespie used the extreme value theory to suggest that one-step beneficial mutations might be expected to exhibit exponentially distributed selective effects (Gillespie, 1983, 1984, 1991). The twenty one-step nystatin adaptive mutations we have acquired here do not immediately appear to fit this prediction, as we recovered an abundance of large-effect mutations (Figure 4.3) whose tolerance to nystatin far exceeds the 4uM exposure concentration they were acquired at. A number of explanations contribute to this finding. Our assay would not have detected small effect mutations, because we required mutations of large enough effect to enable growth in 4uM nystatin. Furthermore, our mutations are not independent of each other; as discussed above, we suspect that the seven different mutations in *ERG6* and eleven unique mutations in *ERG3* are largely loss of function mutations in the ergosterol pathway. In any environment where large effect loss-of-function mutations are available, many different non-synonymous SNPs or indels could be selected, and we might expect these mutations to skew the distribution of beneficial effects toward the maximal fitness effect possible via eliminating the target of selection, here ergosterol. We might also expect that the first mutations acquired have a distribution skewed towards large effect mutations, compared to the distribution of all possible beneficial mutations. Our results are thus more consistent with the extreme value properties in the Weibull domain (where there is a maximal fitness benefit) than in the Gumbel domain used by Gillespie (Joyce *et al.*, 2008).

We found that tolerance across environments frequently exhibited gene-environment interactions, which were typically consistent across different mutations within the same gene.

All BMN lines had a decreased growth rate and decreased biomass production (*OD*48) in the unstressful environment, YPD. Interestingly, we found no evidence that mutations with a larger benefit in nystatin had a greater negative effect in other environments. This is highlighted by a significant positive correlation between IC_{50} in nystatin and growth rate in YPD. When we examined growth in three other stressful environments (copper, ethanol, and salt), we found that beneficial mutations to nystatin had pleiotropic effects that differed substantially among environments (G×E). For example, while all ergosterol mutations examined were beneficial in nystatin, *ERG6* and *ERG7* mutations had reduced tolerance to ethanol, while *ERG3* and *ERG5* mutations were very similar in tolerance to the ancestor. The G×E interactions were so extreme that some mutations exhibited opposite selective effects in some environments. In particular, *ERG6* mutations were less tolerant to copper, while *ERG3*, *ERG5*, and *ERG7* were more tolerant. We call this phenomenon, where two mutations that are beneficial in one environment have selective effects that differ in sign in another environment, "sign G×E" (by analogy to "sign epistasis", Weinreich *et al.* 2005). These experiments demonstrate that although adaptive mutations may show parallel phenotypes in a particular environment (here, in nystatin), effects in other environments of interest can be idiosyncratic and must be specifically examined. For example, although BMN35, with a mutation in *ERG5* shares a similar IC_{50} phenotype with BMN-*erg3* lines in nystatin, ethanol and copper, it has a very different phenotype in salt.

Our results demonstrate that even with a narrow target for adaptation at the gene level (all 35 lines isolated in our screens carried mutations in only four different genes), mutations that appear phenotypically similar in one environment may well present variability in others. Consequently, different subsets of adaptive mutations are likely to be favourable under environmental conditions that require adaptation to more than one selective agent. Furthermore, our results demonstrate that if different first mutations are acquired by separated populations during adaptation to nystatin, this may well place different populations at different locations on the adaptive landscape following shifts in other environmental variables, altering the future evolutionary pathways accessible to these populations. The ability to sequence the entire genomes of multiple adapting lines provides an extremely useful way to explore the range of genetic pathways that evolution can take.

Chapter 5

Mutational effects depend on ploidy level: All else is not equal

5.1 Summary

Ploidy is predicted to influence adaptation directly, yet whether single mutations behave the same in different ploidy backgrounds has not been well studied. It has often been assumed theoretically that aside from dominance, selective parameters do not differ between cells of varying ploidy. We compared the effect size of 20 adaptive mutations in haploids and homozygous diploids and found, surprisingly, that the same mutations often had a much larger effect in haploids than homozygous diploids. This empirical result demonstrates that it can not be assumed that mutations will have the same effect in haploids and homozygous diploids.

5.2 Introduction

The dynamics of evolution are expected to vary between populations that differ in ploidy (Kondrashov and Crow, 1991). In haploid individuals, comprised of a single set of chromosomes, all novel adaptive mutations that arise are immediately "seen" by evolution. The efficacy of selection is strong, as beneficial mutations can be rapidly selected and increase in frequency. In diploids (composed of two sets) or polyploids (multiple sets), mutations generally arise in a single copy which can be partially or completely masked by wild type alleles. The efficacy of selection depends on the dominance properties of the mutation in question (Orr and Otto, 1994). Selection may act quickly in the case of a fully dominant allele, or slowly if the mutation is recessive and must appear in the same genome in multiple

copies before its effects are exposed. This may reduce the rate of adaptation in diploid populations (Chapter 3, Gerstein *et al.* 2011) and may prevent the establishment of recessive or partially recessive adaptive mutations (Anderson *et al.*, 2004).

Whether all else *is* equal between individuals of different ploidy has not been well characterized. Although direct empirical tests are relatively sparse, the existing data suggest ploidy background may well affect the properties of mutations. For example, large chromosomal changes (i.e., rearrangements, deletions, or amplifications) may only be beneficial in diploids (that are able to retain a wildtype copy); indeed, both Gresham *et al.* (2008) and Thompson *et al.* (2006) found the incidence of these mutation types to be higher in evolved diploid lines than haploids. The effect sizes of both deleterious (Szafraniec *et al.*, 2003) and adaptive (Zeyl *et al.*, 2003) mutations have been found to be larger in haploids compared to homozygous diploids in yeast, although this is not always the case (Korona, 1999). The study design of these experiments makes it difficult to interpret potential differences in effect size of single mutations, however, as these studies either compared lines that contained many mutations (Zeyl *et al.*, 2003; Korona, 1999), or conducted a fitness assay where deleterious mutations were assayed in a manner that included the spore germination phase in haploids but not diploids, (Szafraniec *et al.*, 2003). Here we sought to directly assess whether single mutations that confer a benefit to a fungicide, nystatin, have the same effect size in homozygous diploids as in haploids.

5.3 Materials and Methods

(a) ***Mutation acquisition***

Twenty unique mutations that confer tolerance to 4μM nystatin were acquired in haploid lines of *Saccharomyces cerevisiae*, as previously described (Chapter 4). We previously determined that all mutations are in one of four genes that act late in the ergosterol biosynthesis pathway (one mutation in *ERG7* and *ERG5*, seven unique mutations in *ERG6*, and eleven unique mutations in *ERG3*). The majority of mutations are nonsynonymous SNPs or premature stop codons, but the set of mutations also includes three deletions and a duplication. We previously showed that the type of mutation within a gene had no detectable effect and present results here without differentiating between mutation type. Details on the mutations can be found in Table 4.1, which uses the same line designations.

(b) ***Creation of homozygous diploids***

Homozygous diploids were created through a plasmid selection regime that allowed us to create *MATa/MATa* diploids (to avoid the potentially confounding effect of a heterozygous mating locus, Selk and Wills 1998). For each mutation line, two colonies were transformed

using standard protocols (Amberg *et al.*, 2005), with a plasmid containing *LEU2* and one with a plasmid containing *URA3* and *MATα*. Single mutant colonies containing each plasmid were mated and selected on double dropout plates. We then conducted another set of platings to select for plasmid loss. Diploidy was confirmed as previously described Gerstein *et al.* (2011).

(c) ***Nystatin dose-response assay***

We examined the effect of ploidy using a nystatin dose-response assay to measure three parameters: the tolerance of each line to nystatin, measured as the half maximal inhibitory concentration (IC_{50}), the rate of decline in 72 hour culture density at IC_{50} (slope, m_{50}), and the asymptotic level of 72 hour density reached at low levels of nystatin (a). The entire experiment was repeated twice. For each, haploid and diploid mutation lines were struck from freezer stock onto YPD plates and grown for 72 hours. Culture was inoculated into 10mL YPD and grown for 24 hours. We standardized the optical density of all lines and inoculated 10μL into 1mL YPD plus one of ten levels of nystatin (slightly different in each experiment). During each experiment, mutation lines were each grown in four wells of 96-well boxes, which were maintained shaking on a benchtop shaker at 30°C. After 72 hours, wells were mixed and the optical density of 200μL culture was measured on a BioTek plate reader. Data from both experiments were combined for analysis.

The three parameters were fit by maximizing the likelihood of observing the data (Chapter 4). We determined whether haploids and homozygous diploids for each mutation significantly differed by fitting a likelihood model to the combined data, allowing each ploidy to have its own parameter values. We compared the fit of this "full" model to a constrained model with a single value for the parameter of interest (other parameters were allowed to vary), using a likelihood ratio test. If the drop in log-likelihood between the full and constrained model was greater than $\chi^2_{1,0.05}/2 = 1.92$, we rejected the hypothesis that the parameter was the same. We estimated an overall effect of ploidy and gene using the ML parameter estimates from the full model as data in a two-way ANOVA.

(d) ***Growth rate assay***

We also measured the growth rate of all lines at nystatin concentrations below (2μM), at (4μM), and twice the acquisition level (8μM). Growth rate experiments were initiated as above. For each level of nystatin we measured four replicate wells (each contained 150 μL) in non-adjacent wells in the Bioscreen C Microbiological Workstation (Thermo Labsystems). Bioscreen plates were grown at 30°C for 48 hours, with constant shaking; OD readings were taken automatically every 30 minutes. We determined the maximal growth rate for each well as the spline with the highest slope using an analysis program written by Richard FitzJohn in R (R Development Core Team, 2011). We conducted the entire assay

at each level of nystatin on two separate occasions. Significance of the difference in growth rates between haploid and diploid lines was examined for each mutation separately by non-parametric t-tests. The overall effect of ploidy and gene was determined by a two-way ANOVA, as above.

5.4 Results and Discussion

The effect size of nystatin adaptive mutations depended strongly and significantly on ploidy level. Mutations in a haploid background were more tolerant to nystatin than the same mutations in a homozygous diploid background (Figure 5.1). Tolerance (IC_{50}) was significantly affected by gene (F_3 = 221.8, $p < 0.0001$), and ploidy (F_1 = 62.1, $p < 0.0001$), and there was a significant interaction between the two ($F_{3,32}$ = 7.19, p = 0.0008). Haploid lines reached a higher asymptote than diploids (Figure D.1, ploidy: F_1 = 24.4, $p < 0.0001$; gene: F_3 = 15.4, $p < 0.0001$; interaction: $F_{3,32}$ = 1.74, p = 0.19). The slope at IC_{50} was also significant for ploidy and gene (Figure D.1; ploidy: F_1 = 4.5, p = 0.04; gene: F_1 = 3.9, p = 0.018; interaction: $F_{3,32}$ = 0.84, p = 0.48).

Maximal growth rate at three nystatin concentrations was also significantly affected by ploidy background and gene (Figure 5.2). This was not the result of an inherent growth advantage to haploids, as no significant difference was found between wild type haploids and diploids in 2μM nystatin (t_{15}=0.52, p = 0.61, haploid mean: 0.04+/-0.01, diploid mean: 0.03+/-0.01, data not shown). We could not test wild type cells at 4μM or higher levels of nystatin, as by design they were unable to grow without acquiring a mutation. Haploids grew significantly faster than diploids across all lines at 2μM nystatin, and 4μM nystatin (2μM – ploidy: F_1 = 7.8, p = 0.009; gene: F_1 = 101.3, $p < 0.0001$; interaction: $F_{3,32}$ = 0.41, p = 0.75; 4μM –ploidy: F_1 = 6.4, p = 0.017; gene: F_1 = 48.6, $p < 0.0001$; interaction: $F_{3,32}$ = 0.33, p = 0.80). Similar results were obtained when this data was analyzed with a linear mixed-effect model that controlled for batch effects (Table D.1). At 8μM nystatin only lines that carried mutations in *ERG6* (the most tolerant lines, Figure 5.1) were able to grow consistently, yet haploids again grew significantly faster (t_6 = 4, p = 0.01). Lines that carried mutations in *ERG3* and *ERG7* grew stochastically in 8μM nystatin, reminiscent of the growth pattern used to isolate mutation lines (Chapter 4) (Figure D.2). We interpret lines that showed growth to carry novel secondary mutations. Haploids replicates acquired 46 putative mutations, while only 5 diploid replicates showed this pattern (both out of 96 replicates); this difference in putative mutation rate is significant, χ^2_1 = 42.72, $p < 0.0001$. This ploidy-specific difference in mutation acquisition illustrates one of the fundamental differences between lines of varying ploidy. When novel mutations arise in diploid form they are often partially masked by the wild type alleles and thus unable to confer a fitness advantage

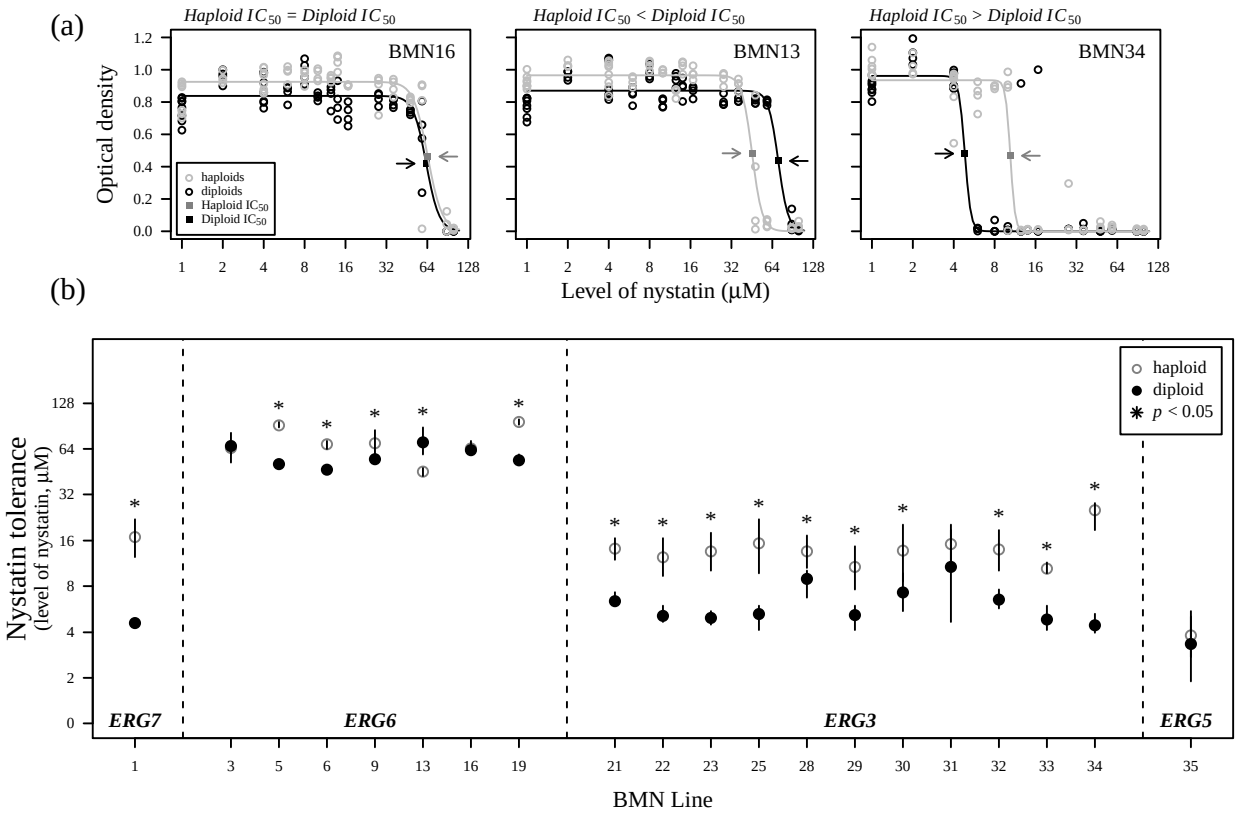


Figure 5.1: Nystatin adaptive mutations generally yield higher tolerance in haploids than homozygous diploids. (a) The dose-response relationship was measured for each adaptive mutation in haploids and homozygous diploids to determine the half-maximal inhibitory concentration of nystatin (indicated by arrows). (b) Significant differences for tolerance were found between ploidy levels and among lines that carried mutations in different genes. Line-specific statistical results are presented in Table D.2.

unless present in more than one copy. We predict that the diploid lines that showed growth contain either rare dominant mutations or a recessive mutation in homozygous form.

Combined, our results demonstrate that haploid yeast carrying adaptive nystatin alleles are more tolerant to nystatin than homozygous diploids. McBride *et al.* (2008) also found that diploid *S. cerevisiae* were more affected than haploids to toxins in their environment. These complimentary results may indicate that physical or expression-level differences exist between ploidy levels of otherwise isogenic yeast that alter the response of cells to either environmental or genetic perturbations. Haploid cells are smaller than diploids cells, and due to their specific geometric shape, haploids have a larger surface area:volume ratio (Weiss *et al.*, 1975). It has long been posited that the fitness of haploid:diploid cells should depend on cell geometry rather than ploidy per se (Weiss *et al.*, 1975). It may be that resistance to nystatin, which acts by binding to ergosterol on the cell surface, is directly correlated to

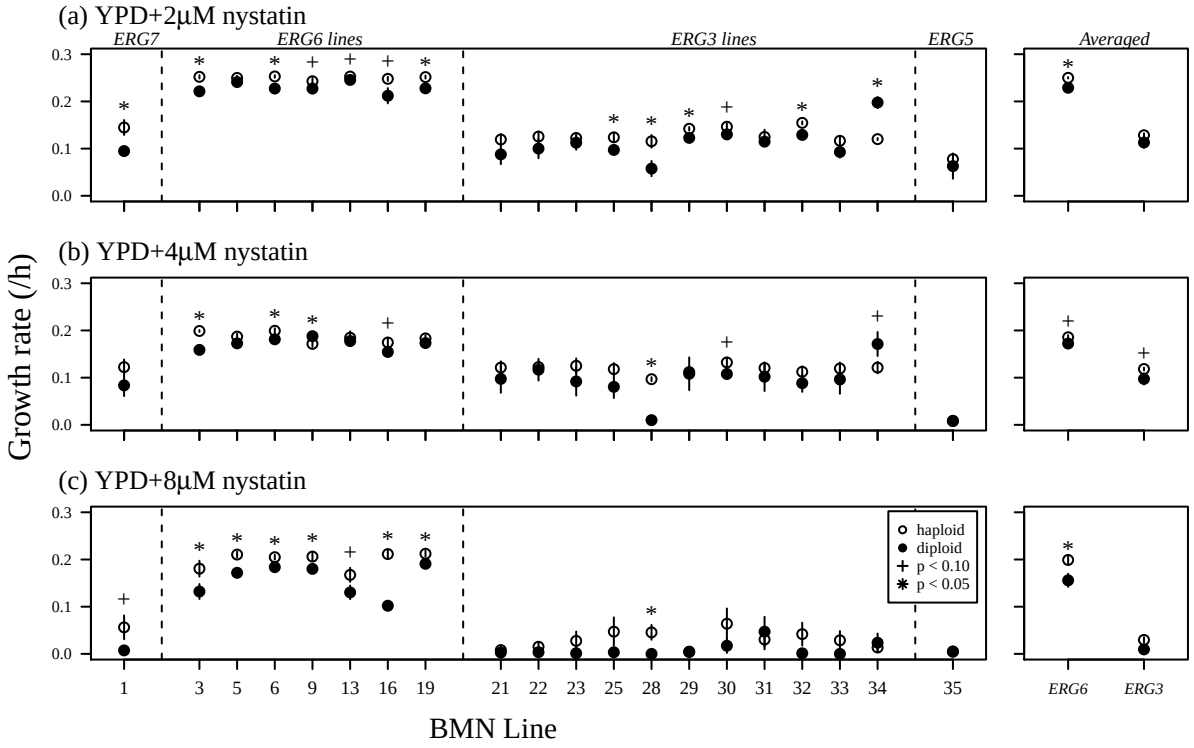


Figure 5.2: Nystatin adaptive mutations tend to have a higher growth rate in haploids than homozygous diploids in (a) YPD + 2μM nystatin, (b) YPD + 4μM nystatin, and (c) YPD + 8μM nystatin. This pattern is also true when averaged across lines that carried different mutations in the same gene (right column, *ERG6* and *ERG3* mutation lines). Note that tolerance in the bioscreen assay cannot be directly compared with tolrance in 1mL deep well plates (Figure 5.1). Line-specific statistical results are presented in Table D.3.

total surface area. Thus, the larger diploid cells could be more affected than haploids by the same concentration of stressor in their environment.

Our results demonstrate that different mutations may well have different effects in haploid and homozygous diploid backgrounds. This is an important result, as theoretical studies often assume that all else is equal when modelling the effect of ploidy, including models about ploidy evolution (e.g., Paquin and Adams 1983; Otto and Goldstein 1992), rates of adaptation (e.g., Orr 2003), and host-parasite interactions (e.g., M'Gonigle and Otto 2011). A better understanding about when (and why) mutational parameters differ between individuals of varying ploidy may help explain why one ploidy level is often favoured over another, a longstanding question in evolutionary biology. Here we have provided one empirical example where the same adaptive mutations yield a greater fitness increase in haploid individuals compared to homozygous diploids.

Chapter 6

Unstable heterozygotes – rapid loss of heterozygosity of adaptive mutations under stress

6.1 Summary

Novel mutations arise in a heterozygous state in diploids, yet much remains unknown about the behaviour of adaptive mutations in heterozygous form. We constructed heterozygous lines of *Sachcaromyces cerevisiae* that contain single adaptive mutations that confer a benefit to nystatin, a fungicide. In an unstressful environment, where the mutations carry a fitness cost in homozygous diploids, we found considerable variation in the dominance of mutations, ranging from dominant (the same phenotype as homozygous mutant diploids) to perfectly recessive (the same phenotype as wildtype diploids). When we examined the mutations in heterozygous form under nystatin stress, we repeatedly found inconsistent growth among replicates within a single growth cycle (24-72 hours). Through targeted Sanger sequencing we uncovered the source of this inconsistency: rapid and repeated loss of heterozygosity (LOH). This result was robust to the initial mutation and the level of stressor in the environment and indicates that heterozygous lines (that do not lose heterozygosity) are no more tolerant to nystatin than the wildtype. Thus, although these mutations have a range of dominance values in an unstressful environment, none of the mutations confer resistance to heterozygotes (i.e., they are all recessive) under the levels of nystatin tested. These results indicate that the dominance properties of beneficial mutations are not consistent across environments. Furthermore, LOH may may play a key role in allowing heterozygous mutations to become homozygous within asexual diploid populations, increasing the rate of adaptation.

6.2 Introduction

Novel mutations typically arise in the heterozygous state in diploids, where their effects are potentially masked by the wildtype copy that remains. For beneficial mutations, this reduces the speed of adaptation, as the full fitness effect of beneficial mutations in heterozygous form is not felt unless they are perfectly dominant (i.e., their fitness is the same as homozygous mutants). If recessive beneficial mutations are common, this provides a strong advantage to haploidy (compared to higher ploidy individuals), as there are no wildtype alleles to mask mutations in haploids, enabling the effects of beneficial mutations to be immediately felt (Orr and Otto, 1994). Similarly, the ability to combine recessive alleles into the same genome is much more easily accomplished through sexual compared to asexual reproduction, a factor that has often been cited as one of the major benefits of sexual reproduction (Fisher, 1930; Kirkpatrick and Jenkins, 1989). An important caveat, however, is that these benefits of haploidy and/or sexuality assume that rare alleles remain heterozygous in asexual diploids. Loss-of-heterozygosity (LOH) has, however, been documented across a range of taxa, and it is plausible that rates of LOH are high enough to mitigate the costs of masking in asexual diploid populations (Mandegar and Otto, 2007). LOH is known to be important in the induction of many cancers, where non-functional tumor-suppressor alleles that are present in populations in heterozygous form, are vulnerable to secondary somatic mutations that inactivate the wildtype allele (e.g., Cavenee *et al.* 1983), thus LOH may also be important from the perspective of beneficial alleles.

Despite the importance of the dominance of mutations to health, adaptation, and evolution, we still lack a complete picture of the phenomenon of dominance and how it depends on the environment and strength of selection. What we do know comes largely from studies of deleterious mutations in single environments, which tend to be partially recessive on average, but with a broad range of dominance coefficients. Studies characterizing mutations in the model species *Drosophila melanogaster*, *Arabidopsis thaliana*, *Caenorhabditis elegans* and *Saccharomyces cerevisiae* generally point to an average dominance coefficient for deleterious alleles of 0.1-0.5 (i.e., a mutation in a heterozygote carries 10-50% of the fitness effect of the same mutation in a homozygote, Crow and Temin, 1964; Mukai *et al.*, 1972; Simmons and Crow, 1977; Houle *et al.*, 1997; Garcia-Dorado and Caballero, 2000; Vassilieva *et al.*, 2000; Zeyl and DeVisser, 2001; Chavarrias *et al.*, 2001; Szafraniec *et al.*, 2001; Fry, 2004). The same conclusion is reached through studies subjecting *S. cerevisiae* or *C. elegans* to EMS (Peters *et al.*, 2003; Szafraniec *et al.*, 2003). The yeast gene deletion projects, where every gene in the genome has been deleted, indicates that fewer than 10% of heterozygous gene deletions have a clear fitness detriment in rich medium (YPD) in either *S. cerevisiae* (Winzeler, 1999) or *Schizosaccharomyces pombe* (Kim *et al.*, 2010). An even lower number of loci (0.4%) appear to affect fitness when one copy is deleted in *Drosophila* (Lindsley *et al.*,

1972).

The dominance of beneficial mutations has been much less characterized, yet it is precisely these mutations that enable adaptation. Adaptation can occur from either standing genetic variation or via the incorporation of novel beneficial mutations (Hermisson and Pennings, 2005). New beneficial mutations establish in populations with a probability that is proportional to their effect size (s) scaled by their dominance coefficient (h). Thus, if all else were equal, new dominant mutations would be more likely than recessive ones to contribute to adaptation ("Haldane's sieve", Haldane 1927). However, if adaptation occurs via alleles already present, it has been shown theoretically that the probability of fixation for newly favourable alleles (which were formerly deleterious or neutral) depends less on the dominance of mutations (Orr and Betancourt, 2001; Hermisson and Pennings, 2005). In fact, if the dominance of beneficial alleles is positively correlated with their dominance in a previous environment where they were deleterious, then Haldane's sieve vanishes. Whether there is such a correlation, however, has been untested. An empirical examination of the dominance characteristics of beneficial mutations requires a set of beneficial mutations that is not biased by any potential effects of Haldane's sieve.

In this study, we aimed to measure heterozygous fitness for a large number of beneficial mutations previously isolated from haploid *S. cerevisiae*. These mutations were acquired in haploids, irrespective of their dominance in diploids. Specifically, we characterized twenty unique mutations that confer a large fitness benefit in YPD+4μM nystatin in haploid lines of *S. cerevisiae* (Chapter 4). These mutations bear a fitness cost relative to wildtype in YPD, an unstressful environment (Figure 4.4). Thus, these same mutations can be either beneficial or deleterious, depending on the environment in which they find themselves. The genetic basis of each mutation was identified through whole genome sequencing (WGS); each line carries a mutation in one of four genes that act late in the ergosterol biosynthesis pathway (one mutation in *ERG7* and *ERG5*, seven unique mutations in *ERG6*, and eleven unique mutations in *ERG3*), with very few to no additional mutations throughout the genome. As ergosterol is the primary target of nystatin (Woods, 1971), we postulate that the mutations function to alter the ergosterol content of the cell membrane, and thus reduce the ability of nystatin to bind and form detrimental membrane pores. The majority of our mutations are nonsynonymous SNPs or premature stop codons, but the set of mutations also includes a small deletion in *ERG6*, a small deletion and a large duplication in *ERG3*, and a large deletion in *ERG5* (Table 4.1). Each mutation was previously shown to confer resistance to nystatin in homozygous diploids (Chapter 5).

Here we compared diploid heterozygous lines to both wildtype diploids and homozygous mutant diploids. We assayed heterozygous fitness for mutations known, in the haploid state, to be beneficial in one environment (nystatin) and detrimental in another (YPD, an unstressful environment), to determine whether dominance was consistent across environ-

ments. If this were the case, the limitations of Haldane's sieve would be weakened, and beneficial mutations with a much larger range of dominance values could contribute to evolution.

6.3 Results

6.3.1 *Heterozygotes grow better than homozygotes in an unstressful environment*

By crossing wildtype haploids and haploids containing previously characterized nystatin beneficial mutations (Chapter 4), we created heterozygous and homozygous mutant diploids. We first compared the fitness of these diploids to wildtype diploids in an unstressful environment, the standard rich medium YPD. We measured the maximal growth rate ("growth rate") over 24 hours and the optical density at 24 hours ("biomass production") using the automated Bioscreen C Microbiological Workstation (Thermo Labsystems). If mutations were perfectly recessive, we would expect heterozygotes to grow at the same rate as wildtype diploids, while if heterozygotes were perfectly dominant, they would behave the same as mutant homozygotes. As shown in Figure 6.1, heterozygotes generally have a faster growth rate and reach higher biomass production in YPD than homozygous mutants.

We conducted the entire experiment on two separate occasions and though results were generally very consistent among the four replicates done for each line within an experiment, we found considerable variation between the two experiments; the error bars in Figure 6.1 thus reflect the error in trait means between the two replicate experiments, though statistical tests were conducted on all replicates. Interestingly, heterozygous replicates showed considerably more variation both within and between experiments, a result that we repeatedly found under varying assays and conditions (see below).

All mutations in homozygous form grew significantly slower and reached significantly lower biomass than did wildtype diploids (Table E.1). Similarly, almost all homozygous mutants grew significantly slower than did heterozygous lines (Table E.2) and exhibited lower biomass production (Table E.3). The only exceptions were BMN35 (the only line with a mutation in *ERG5*, which grew the same in homozygous and heterozygous backgrounds for both traits) and BMN6 (which has a mutation in *ERG6* and did not significantly differ in biomass production); these lines seemed to contain mutations that are dominant in YPD for these traits. Comparing heterozygotes to wildtype diploids, we found that only eleven lines were significantly different for at least one trait (maximal growth rate: Table E.2, biomass production: Table E.3). In the remaining cases, the variation between replicates may have reduced our power to detect small differences, and it seems unlikely (though possible) that all other lines are perfectly recessive. Nevertheless, we observed a large range of dominance coefficients for the mutations in these lines (Table 6.1).

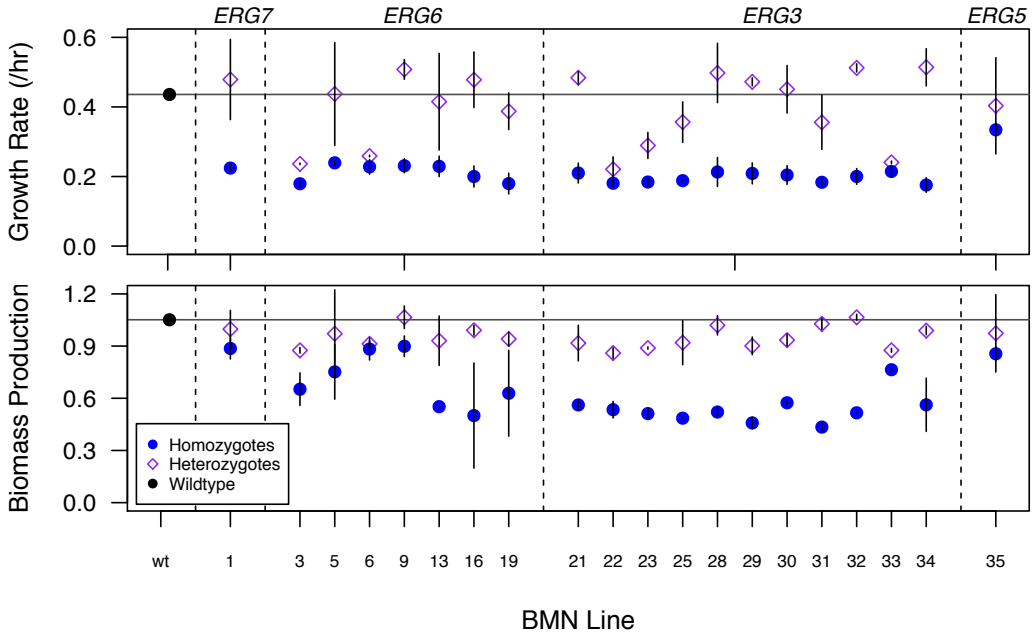


Figure 6.1: Lines with heterozygotes mutations (purple open diamonds) perform better than lines with homozygotes mutations (blue closed circles) in an unstressful environment (YPD). Heterozygotes generally grow faster (top) and are better able to convert nutrients into cellular matter (biomass production, bottom) compared to homozygotes. Lines on the x-axis are arranged by the gene that carries a mutation; the lines farther to the left carry mutations in genes that are most upstream from the production of ergosterol, the end product of the pathway. When multiple mutations are present within a gene, the lines are arranged from mutations earlier to later in the gene. Error bars reflect the standard error of the means for each line from experiments conducted on two separate days.

Table 6.1: Estimated dominance coefficients in YPD for heterozygous lines that grew significantly different than wildtype. Standard error is the error of the means from two replicate experiments. Note that negative numbers imply overdominant mutations (i.e., the fitness of heterozygotes is above that of mutant and wildtype homozygotes). Dominance calculated as
$(fitness_{heterozygote} - fitness_{wildtype})/(fitness_{homozygote} - fitness_{wildtype})$

		Dominance estimate (h)	
BMN Line	Gene	Growth rate	Biomass production
1	*ERG7*	-0.20±0.38	0.32±0.72
3	*ERG6*	0.78±0.38	0.44±0.02
5	*ERG6*	0±0.6	0.27±1.14
6	*ERG6*	0.85±0.12	0.82±0.49
9	*ERG6*	-0.35±0.03	-0.09±1.24
13	*ERG6*	0.10±0.53	0.24±0.20
16	*ERG6*	-0.18±0.21	0.11±0.02
19	*ERG6*	0.19±0.14	0.26±0.20
21	*ERG3*	-0.21±0.06	0.27±0.14
22	*ERG3*	0.84±0.13	0.37±0.16
23	*ERG3*	0.58±0.13	0.30±0.10
25	*ERG3*	0.32±0.15	0.23±0.15
28	*ERG3*	-0.30±0.08	0.13±0.09
29	*ERG3*	-0.28±0.26	0.06±0.01
30	*ERG3*	-0.16±0.18	0.25±0.17
31	*ERG3*	-0.06±0.16	0.25±0.01
32	*ERG3*	0.32±0.38	0.04±0.14
33	*ERG3*	-0.32±0.19	-0.03±0.08
34	*ERG3*	0.88±0.03	0.61±0.08
35	*ERG5*	0.32±1.1	0.40±0.89

6.3.2 *Heterozygotes grow stochastically relative to homozygotes under stressful conditions*

We next compared the growth profile of homozygous and heterozygous mutation lines grown in nystatin, the stressor in which the mutations were acquired. Specifically, we looked at growth at the level of nystatin used to acquire mutations (YPD+4μM nystatin) and at half this level (YPD+2μM nystatin) for 72 hours in the bioscreen (see Methods). Through an examination of the raw growth curves we found that heterozygous replicates from all lines exhibited stochastic growth in both levels of nystatin (Figure 6.2), reminiscent of the growth patterns initially observed when haploids first acquired mutations (Chapter 4). We subsequently repeated this experiment and found that stochastic growth was present for heterozygous lines in each case. About half of heterozygous well replicates grew within 72 hours in YPD+2μM nystatin (we observed growth above the background in 365 replicate wells out of 740), while significantly fewer replicate heterozygous wells grew in YPD+4μM

nystatin (249 of 630 replicates, p = 0.0003, by a Fisher's Exact Test). We suspected that heritable genetic changes had arisen in these lines; when culture from wells that showed growth was placed back into the stressful medium, growth trajectories were much more consistent, similar to those of the homozygous lines (Figure E.1). We subsequently froze down culture from replicate wells that showed growth in the second bioscreen experiment for future use.

We then wished to determine the level of nystatin at which heterozygotes begin to grow stochastically. We had previously found that homozygous diploid lines were able to grow consistently in YPD + 4μM nystatin in deep well boxes (Chapter 5), and we thus focused on growth in nystatin at or below this level. Homozygous and heterozygous lines were each grown in six replicate deep wells containing 1mL YPD and one of five nystatin concentrations (0.5μM, 1.0μM, 1.5μM, 2.0μM, and 4.0μM). We also assayed replicates from five wildtype homozygous colonies. After 72 hours of growth, we measured the optical density of each well. Surprisingly, although in many cases the mean of heterozygous replicates appeared to be intermediate between homozygous mutant diploids and wildtype diploids, the majority of heterozygous replicate wells either grew fully or not at all (Figure 6.3). Interestingly, although wildtype replicates grew consistently in both YPD+0.5μM and YPD+1.0μM nystatin, heterozygous lines grew consistently only in 0.5μM nystatin and continued to show inconsistent and stochastic growth at all levels above this. This pattern suggest underdominance at 1μM nystatin, with both wildtype and homozygotes growing faster, on average, than heterozygotes at low levels of nystatin.

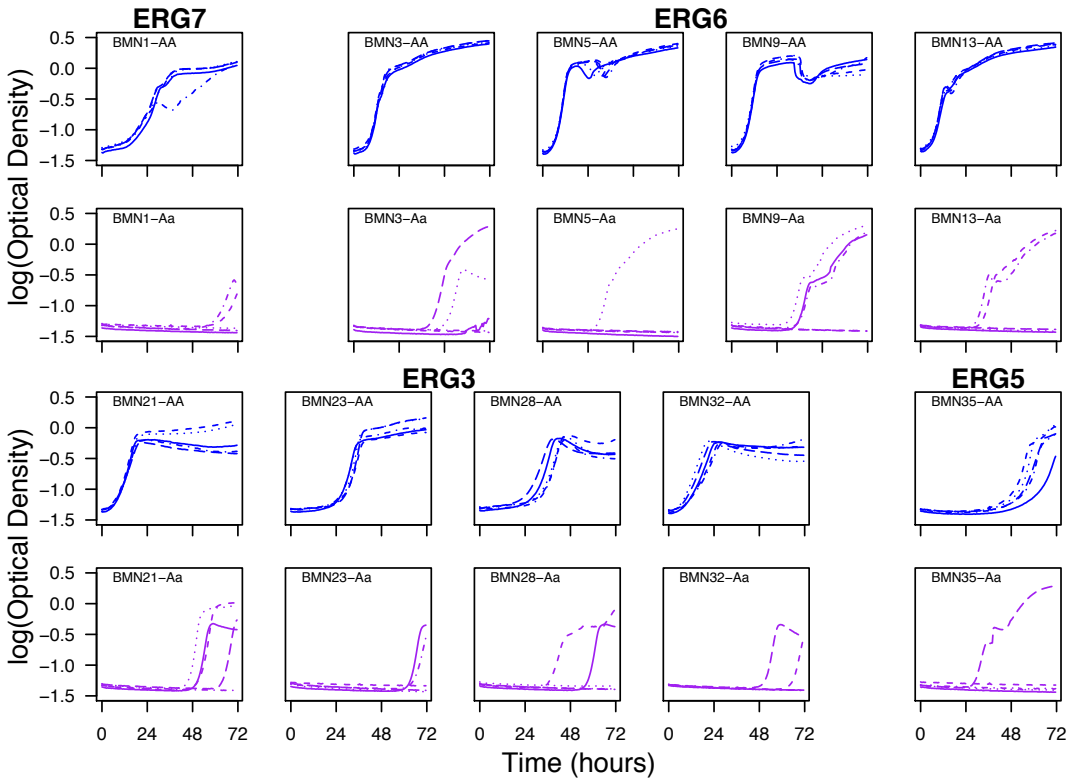


Figure 6.2: Homozygous diploids (lines shown in blue, top panel of each pair) and heterozygous lines (purple, bottom panels) were grown in YPD+2μM nystatin. Depicted is a representative subset of the 20 mutation lines, each grown in five replicate wells (different dashed curves). The pattern of growth is very similar in YPD+4μM nystatin. All replicate wells for each mutation line were initiated with the same culture. Note that the five replicate growth curves are similar in homozygous mutant diploids (blue) but not in heterozygotes (purple).

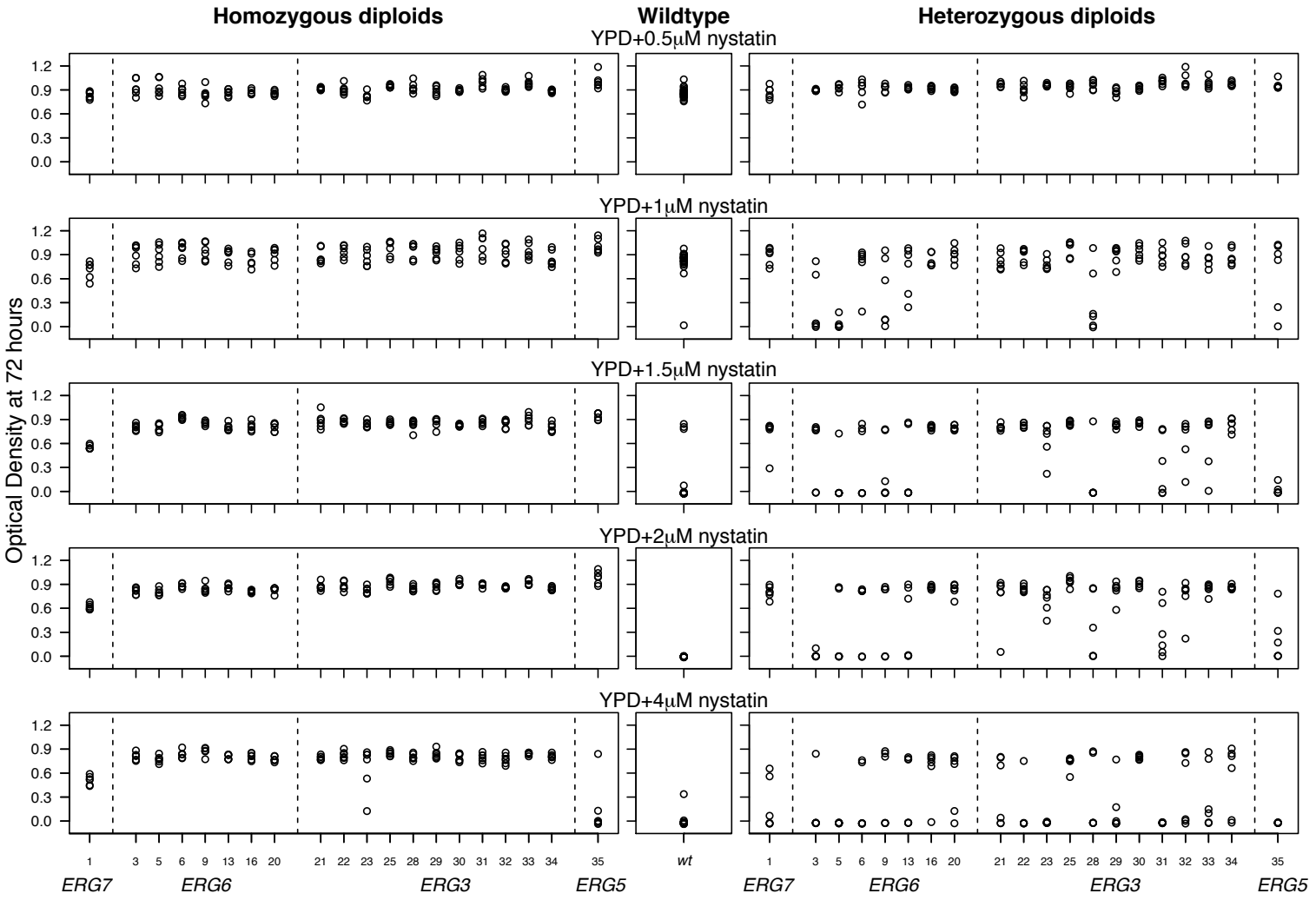


FIGURE 6.3: Heterozyous replicates grew stochastically in deep well boxes at levels of nystatin above 0.5μM. Four replicates were measured after growth over 72 hours for homozygotes (left) and heterozygotes (right) from each line, except the wildtype (center), with 8 replicates.

6.3.3 *Loss of heterozygosity explains stochastic heterozygous growth*

Given the above results, we hypothesized that heritable genetic changes were occurring in heterozygotes, through the acquisition of secondary mutations, loss of heterozygosity of the initial mutation, or allelic silencing of wildtype alleles. We focused our effort on determining whether the heterozygous replicates that showed growth (hereafter referred to as "het-grow" lines) had retained heterozygosity at the initial mutation locus. In total we examined 78 replicates from across multiple lines isolated and frozen down immediately after growth in YPD, YPD+1μM nystatin, YPD+2μM and YPD+4μM nystatin (i.e., 72 hours after replicate wells were inoculated with culture from heterozygous lines; see Tables E.4-E.7 for full information, including colony designations). For each het-grow line we isolated culture from freezer stocks and obtained genomic DNA using standard methods. Het-grow lines isolated from YPD+2μM nystatin were streaked onto plates to colony level, while het-grow lines from YPD, YPD+1μM nystatin and YPD+4μM nystatin were sequenced as population samples, which allowed us to assess levels of polymorphism. We PCR amplified $\sim$ 1000bp around the known mutation locus, and Sanger sequenced the resulting fragments.

Whether or not heterozygosity was retained for the initial mutation depended on the environment from which they were isolated (Figure 6.4). All het-grow lines isolated after growth in YPD remained heterozygous, indicating that the increased variation we found within heterozygous replicates from the same line (relative to homozygotes) in YPD (Figure 6.1) was not due to genetic changes at the initial mutation locus (Figure 6.4). In contrast, the majority of replicates isolated after growth under nystatin stress were homozygous for the initial mutation (Figure 6.4). In YPD+1μM nystatin, 23 of 28 replicates were homozygous for the initial mutation. Of the remaining five replicates, we sequenced only one replicate population (BMN13-10E), that retained full heterozygosity. The four population replicates initiated from BMN35 culture, which carried a 60bp deletion in *ERG5*, were polymorphic for heterozygous and homozygous mutant diploid individuals (Figure E.2).

None of the 37 het-grow lines isolated after growth in YPD+2μM or 4μM nystatin remained heterozygous for the initial (ergosterol) mutations (Figure 6.4). The majority of het-grow lines (29/37) had become homozygous for the ergosterol mutations, though we found eight replicates that were homozygous wildtype at the initial locus (Table E.8). These eight lines were all consistent with contamination events during the set of experiments used to isolate the het-grow lines. We Sanger sequenced *ERG6* and *ERG3* in these putative contaminant lines; all lines contained ergosterol mutations carried by other lines in our screen (Table E.8). The potential sources of contamination are different in the bioscreen and deep well experiments. In the bioscreen experiments (where we found 6 contaminated lines), homozygous culture was grown in wells directly beside heterozygous culture, thus it is possible that some het-grow lines were contaminated from adjacent homozygous culture. In the

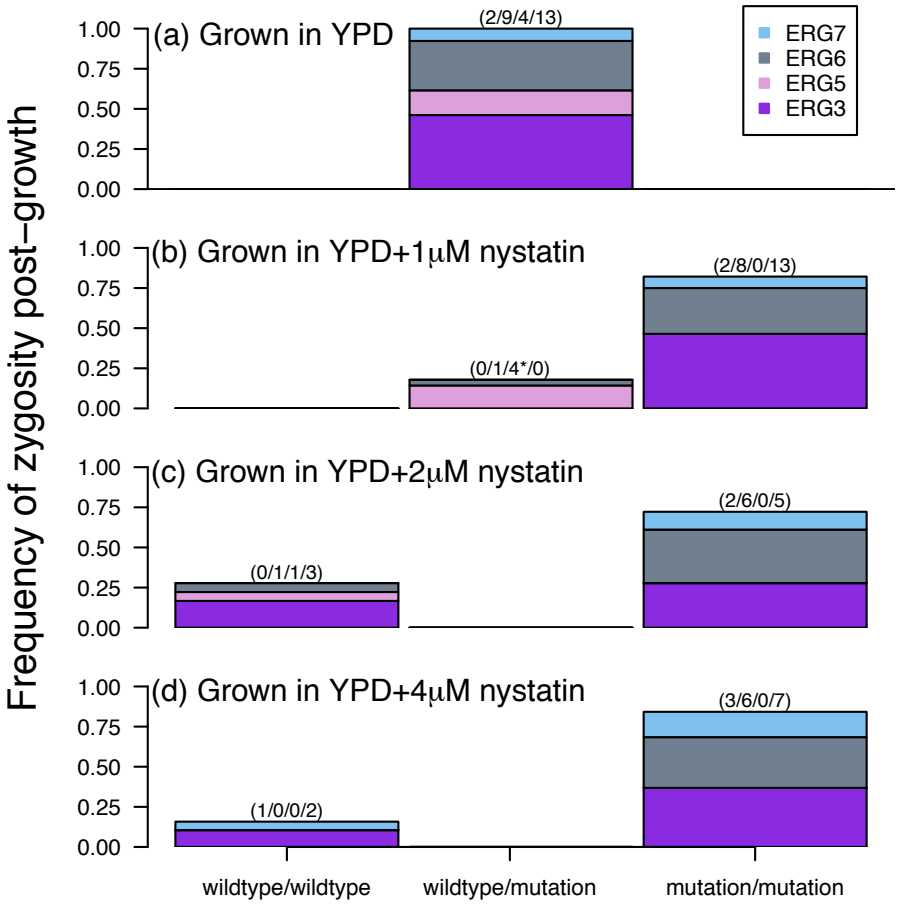


Figure 6.4: Adaptive mutations in heterozygous form rapidly lose heterozygosity to enable growth under stress. The numbers above the bars indicate the number of Sanger sequenced lines that carry mutations in the four genes (*ERG7*/*ERG6*/*ERG5*/*ERG3*). Wildtype/wildtype * indicates *ERG5* populations that appear polymorphic for heterozygous and homozygous mutant individuals.

deep well experiments, however, homozygotes and heterozygotes were grown in different boxes, limiting simple contamination (without LOH) as a source of homozygotes. Nevertheless, two cases of homozygosity for the wildtype allele were observed in the YPD+4μM nystatin deep well boxes. In each of these cases, there was evidence that a neighbouring well carrying the contaminating ergosterol mutations (Table E.8) had undergone LOH and then contaminated the well in question during sampling (boxes were sampled at multiple timepoints to characterize stochastic growth). We did not include the contaminated lines in the remainder of our analyses.

We next considered whether the lines that appeared to be homozygous for their initial ergosterol mutation and that were interpreted as LOH events might also represent 'cryptic' contamination (i.e., contamination from a homozygous diploid that carries the same mutation). In the bioscreen experiment, a number of putative het-grow lines (9) carried secondary (non-ergosterol) mutations (Table E.9). Through Sanger sequencing, we found that all lines remained heterozygous for the secondary mutations, arguing against contamination from diploids that were homozygous at the start of the experiment (which would have been homozygous for all these mutations as well). Thus, in no case where LOH was inferred for the original ergosterol mutation was there evidence of contamination.

6.3.4 *Investigating the mechanism by which heterozygosity was lost*

LOH can be accomplished by a number of mechanisms, including gene conversion, mitotic recombination, and copy number changes. We first investigated copy number changes in het-grow lines to see if the initially heterozygous lines had become haploid (whole genome LOH), aneuploid (chromosomal LOH), or if gene copies had been locally deleted (local LOH). We used flow cytometry to measure the ploidy of all 70 het-grow lines and found only a single ploidy change: BMN3-188 became haploid upon growth in 2μM nystatin. To identify partial or complete aneuploidy, we selected six lines (including BMN3-188) isolated after growth in YPD+2μM nystatin that exhibited the lowest G1 mean from flow cytometry (potentially implying whole or partial aneuploidy, data not shown). We whole-genome resequenced these lines using an Illumina Hi-Seq 2000. From analyses of the depth of coverage across the genome, we saw very few cases of whole-chromosome aneuploidy (none inolving the chromosomes that the ergostergol genes are located on, Figure E.3) and no cases of large scale partial aneuploidy in the six lines (Figure E.4). The haploid replicate BMN3-188 was found to carry additional copies of chromosomes 2 and 9, while one line bearing an *ERG3* mutation, BMN23-144, was found to have lost one copy of chromosome 3 (which is not the chromosome on which *ERG3* resides, which is chromosome 12).

Partial aneuploidy of local gene deletion could also enable LOH. We next investigated changes in copy number at the level of each ergosterol gene. To avoid ambiguities due to alignment issues had there been gene loss, we directly counted the number of 100bp fragments in the unaligned fastq files obtained by Illumina sequencing that contained the first or last 18bp of each ergosterol gene (*ERG7*, *ERG6*, *ERG3*, and *ERG5*, using "grep" to sum counts in the forward and reverse direction and averaging over the total from the beginning and end of the gene). As this is the bioinformatics equivalent to a qPCR analysis of WGS data, we refer to this technique as *in silico* qPCR. *in silico* qPCR showed no evidence of localized gene duplication or loss in any of the ergosterol genes or four control genes in any of the lines (Figure E.5).

The mitotic repair of double strand breaks by homologous recombination has previously been found to be the most common mechanism enabling loss of heterozygosity in *S. cerevisiae* (St Charles *et al.*, 2012; Hiraoka *et al.*, 2000, and references within). As the design of our experiment focused on lines that were identical except for newly arisen beneficial mutations, our strains were not constructed with markers that would allow us to quantify the crossover events or gene conversion patterns that contributed to LOH. As with our contamination check, however, we examined the secondary mutations that were present across the genome in some of our lines (Table E.9, Table E.10). Sanger sequencing these secondary mutations indicated that in all cases the secondary mutations were heterozygous, indicating that LOH was a local event.

The rate of mitotic recombination has been shown to increase linearly with distance from the centromere in *S. cerevisiae* (Mandegar and Otto, 2007) and *Candida albicans* (Forche *et al.*, 2011). LOH events have, however, been recorded in yeast regions located very close to a centromere (Minet *et al.*, 1980; Hiraoka *et al.*, 2000; St Charles *et al.*, 2012). The four genes we identified in our screen are located at variable distances from their centromeres: *ERG5* is the furthest from the centromere ($\sim$195kb away), while *ERG6* is very close (only $\sim$15kb, with a genetic distance of 2cM, Saccharomyces Genome Database), compared to $\sim$134kp for *ERG7* and $\sim$102kb for *ERG3*. If the rate of mitotic recombination is the primary driver of LOH in our lines we would thus predict lines with a mutation in *ERG6* to undergo LOH less frequently than lines with mutations in other genes. This was not the case (Figure 6.5). Replicates with mutations in *ERG6* seemed to LOH at a very similar rate to mutations in other genes, however, while very few replicates initiated from BMN35, with a mutation in *ERG5*, the gene that is furthest from a centromere, exhibited growth in YPD+2μM nystatin or YPD+4μM nystatin (Figure 6.5). This suggests that mitotic recombination is not the primary mechanism of LOH. It may be that other mechanisms, such as whole-chromosome loss of heterozygosity (e.g., due to segregation of homologues rather than sister chromatids during mitosis), also occur at an appreciable frequency. A recent experiment in *C. albicans* lines found whole chromosome loss of heterozygosity to be the most frequent mechanism

enabling LOH in lines subjected to fluconazole, an antifungal drug that is similar to nystatin (Forche *et al.*, 2011).

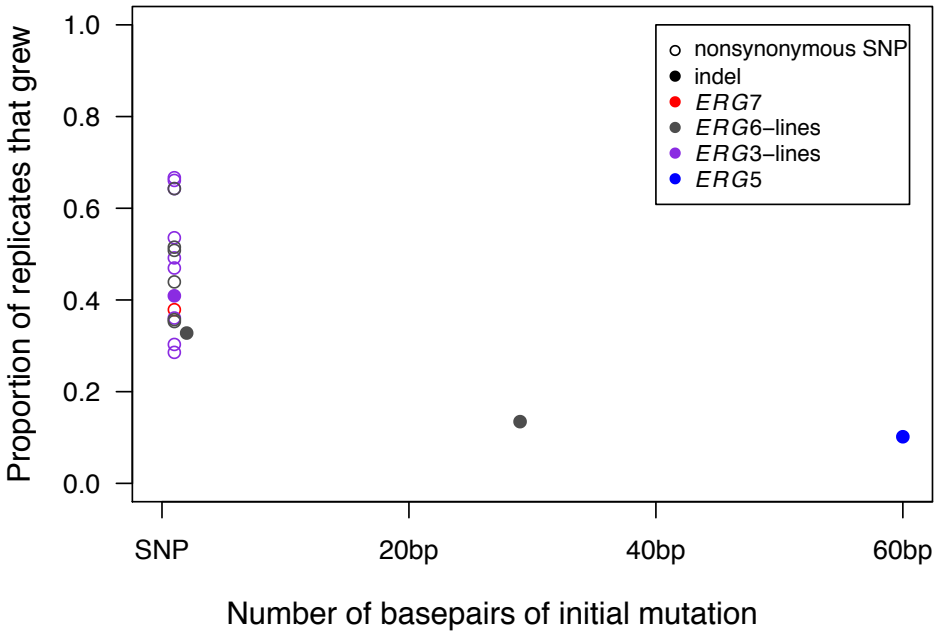


Figure 6.5: We found variation in the number of replicates from each line that grew in YPD+2μM nystatin and YPD+4μM nystatin. This figure represents all replicates from each heterozygous line, grown in both bioscreen and deep well experiments at both levels of nystatin.

Another possible explanation for LOH is gene conversion. BMN35 carries a 60bp deletion in *ERG5*, in contrast to the majority of mutation lines that carried nonsynonymous SNPs, and so it may be less likely to undergo homology-dependent gene conversion. To investigate whether the type of mutation (i.e., indel versus nonsynonymous SNP) might have affected the rate of LOH, we examined the other three mutation lines that carried indels. BMN29, which carries a 29bp duplication in *ERG3* shares a similarly low rate of growth (only 7 of 52 BMN29 replicates grew, compared to 7 of 61 BMN35 replicates). As shown in Figure 6.5, the other two mutation lines (BMN13, with a 2bp deletion in *ERG6* and BMN34 with a 1bp deletion in *ERG3*) are also among the lines with the lowest number of replicates that grew. Without a larger dataset we do not wish to over-generalize our results, beyond noting the trend that insertions and deletions in our mutation set seem to undergo LOH much less readily than nonsynonymous SNPs.

6.4 Discussion

Our results indicate that the dominance properties of mutations are not constant among environments. The mutations we examined are deleterious in an unstressful environment in both haploids and homozygous diploids, and exhibited a wide range of effects in heterozygous form, ranging from perfectly recessive to perfectly dominant (Table 6.1). For the majority of mutations, we found that heterozygotes did not bear the same fitness cost as homozygotes in an unstressful environment, implying that they would not be as strongly selected against in this environment. Under nystatin stress where the mutations are beneficial, however, our results indicate that the heterozygous lines that do not lose heterozygosity are no more tolerant to nystatin than the wildtype (Figure 6.3), indicating that these mutations are recessive in the levels of nystatin we examined. Surprisingly, we even observed LOH at 1μM nystatin, a level of stressor where the wildtype appears capable of growth (Figure 6.3). This result is consistent with either underdominance of heterozygotes (i.e., the heterozygotes are less fit than wildtype cells), or a very high beneficial mutation rate allowing all wildtype replicates to grow. Because these wildtype cells are diploid, if wildtype growth at 1μM nystatin is due to new mutations, these mutations must either be a different set of dominant mutations than those that we have previously sampled, or new mutations coupled with extremely rapid loss of heterozygosity.

Although a number of studies have characterized in depth the dominance of deleterious mutations, to our knowledge no study has systematically characterized the dominance of a large number of mutations that are beneficial in one environment, nor examined the same beneficial mutation in multiple environments. These questions deserve considerably more attention in the future. The striking variation in dominance coefficients across environments that we observe could be explained in at least two ways. From the perspective of a single pathway, sensitivity to alterations in gene product may well depend on the environment; for example, fitness may be robust to small changes in protein product in a non-stressful environment (mutations being recessive) but become sensitive in a stressful environment (mutations being dominant). This reasoning is in line with metabolic theory (Kacser and Burns, 1981) as well as theories about the interplay between genetic and environmental robustness (Meiklejohn and Hartl, 2002). Secondly, changes in dominance across environments might be explained on the basis of the pleiotropic effects of genes, e.g., if the mutation has pleiotropic effects that are not felt in all environments. Regardless of the explanation, idiosyncratic dominance relationships across environments make it clear that a more nuanced form of Haldane's sieve must act upon beneficial mutations. Specifically, if adaptation occurs from formerly deleterious mutations maintained as standing genetic variation, the subset of mutations of this class that are most likely to contribute will be those that are recessive in the previous environment (and so are at relatively high frequency) but

dominant in the novel environment (so that each copy is more likely to fix).

If adaptive mutations frequently undergo LOH, the rate of adaptation of asexual diploids may be much faster than predicated theoretically. To the extent that beneficial mutations must remain heterozygous in the absence of sex, the rate of adaptation of asexuals would be reduced relative to that of sexual species (Kirkpatrick and Jenkins, 1989), but LOH can minimize or eliminate this difference (Mandegar and Otto, 2007). Similarly, if heterozygous mutations are generally recessive, frequent LOH could reduce differences in the rate of adaptation of diploids compared to haploids. Indeed, though few studies have directly quantified (or looked for) LOH in evolutionary experiments, it is a regular phenomenon in *S. cerevisiae*. Gresham *et al.* (2008) noted that a nonsynonymous SNP in *GSH1* that appeared in less than 10% of diploid individuals at generation 100 of a continuous culture experiment was found to be homozygous in at all individuals by 150 generations of evolution. Similarly, Anderson *et al.* (2004) observed that diploids evolved at high fluconazole acquired resistance much slower than haploids (consistent with recessive mutations); when they sporulated the diploids resistance for fluconazole did not segregate, consistent with a LOH event.

Our findings of widespread LOH even within 72-hour fitness assays indicate potential difficulties in accurately measuring the fitness of heterozygotes, particularly if researchers do not account for the possibility of LOH. Our experiments were all conducted at a very short timescale, and indicate that LOH can be both rapid and frequent, particularly with large population sizes and strong selection. In YPD+1μM nystatin, for example, many lines exhibited full growth by 72 hours Figure 6.3, suggesting that these mutations are dominant. It was only through sampling at multiple time points (which revealed stochastic growth) and post-growth sequencing that the truth emerged: these lines were no longer heterozygous.

6.5 Conclusions

Our results indicate that the dominance of adaptive mutations may depend on the environment. We assayed 20 mutations in heterozygous form in an unstressful environment, and found a wide range of dominance values among them, ranging from wholly dominant to wholly recessive. In the stressful environment where these mutations confer a fitness benefit to homozygous diploids, however, none of the heterozygotes were able to grow in a level of stressor beyond that which the wildtype was able to grow, indicating that all mutations were wholly recessive. Upon exposure to stressor, however, heterozygotes were able to rapidly lose heterozygosity, exposing the adaptive mutations. Although the rate of LOH per cell cycle might be low, if the LOH products are strongly selected, the frequency

of genotypes that arose because of LOH can become high within a very short number of generations. As new mutations arise in heterozygous form in diploids, it has been hypothesized that natural selection will act by selecting beneficial mutations that are predominantly dominant, a phenomenon referred to as 'Haldane's sieve. Our results suggest that spontaneous LOH may frequently act in diploid genomes, however, and allow recessive beneficial mutations to "see the light" of natural selection.

6.6 Materials and Methods

(a) ***Mutation acquisition***

We created heterozygous diploids from the haploid BMN lines examined in Chapter Four. A single colony from each mutation line was transformed with a plasmid containing LEU2 while an ancestral colony (in background BY4741, *MATa his3*$\triangle$1*leu*2$\triangle$0*met*15$\triangle$0*ura*3$\triangle$*o*, derived from S288c) was transformed with a plasmid containing URA3 and MATα. Single transformed colonies of each line (mutant and wildtype) were acquired on the appropriate dropout plates, co-inoculated into 200μL YPD, and spotted onto YPD plates. We then replica plated from YPD onto -ura, -leu and YPD plates; a single colony that grew only on YPD plates was selected as the final mutant heterozygous diploid for each line. Each mutant diploid was assayed for ploidy as a double check that the matings were conducted properly, as previously described (Gerstein and Otto, 2011). We Sanger sequenced half the lines to confirm that the mutation locus in each line was heterozygous. Homozygous diploid lines were obtained in the same fashion.

6.6.1 *Bioscreen experiments*

We compared the growth curves of heterozygotes and homozygotes in YPD, YPD+2μM nystatin and YPD+4μM nystatin using a Bioscreen C Microbiological Workstation (Thermo Labsystems). Each bioscreen experiment was initiated in the same fashion. All mutation lines were streaked onto plates from culture maintained at -80°C onto a YPD plate and allowed to grow for 72 hours. We then picked a single colony and placed it into 10mL YPD, allowing it to grow overnight. We standardized the optical density of all lines to that of the lowest optical density. 10μL of this standardized culture was then inoculated into 1mL of the environment that was being tested. 150μL of diluted culture from each line of interest was then placed into 4 replicate non-adjacent bioscreen wells. To measure maximal growth rate in YPD, we measured the maximal growth rate through a spline-fitting procedure described previously (Gerstein and Otto, 2009). We ran two bioscreen experiments in YPD+2μM nystatin and YPD+4μM nystatin. Het-grow lines were only isolated

after the second experiment (as we uncovered the pattern of stochastic growth after the first experiment, but did not freeze down culture). After 72 hours in the second bioscreen we examined the growth curves. For heterozygous wells that showed growth after the YPD+2μM bioscreen experiment, we streaked out frozen culture to single colony on YPD plates and haphazardly froze a single colony in an eppendorf in 15% glycerol. For the wells that showed growth after the YPD+4μM experiment, we added 150μL 30% glycerol, mixed well, pipette into eppendorfs, and froze the entire culture.

6.6.2 *Deep-well experiments*

All deep well experiments were initiated in a similar fashion as the bioscreen experiments. We grew all 20 unique mutations lines, in both homozygous and heterozygous states, plus the wildtype control (BY4741) from frozen stock on YPD plates. We haphazardly chose a single colony from each mutation line and five colonies from the wildtype, placed each colony in 10mL YPD, and grew them overnight. We then standardized 200μL from each line to the lowest optical density. We added 12μL of standardized culture to 600μL of medium of containing the appropriate stressor and placed 1mL of the diluted culture into four replicate deep wells. Heterozygous and homozygous replicates were grown in separate boxes; replicates from the same line were never placed beside each other.

As with the bioscreen experiments, we conducted the deep-well box experiments twice, yet only froze culture from the second experiment, In the first experiment we assayed culture only at 72 hours. In the second, we assayed the optical density of all boxes every 24 hours (data not shown). We manually mixed all wells and measured 150μL aliquots on a BioTek microplate reader (BIoTeck Instruments, Inc. Winooski, VT). Boxes containing heterozygous replicates grown in YPD+1μM nystatin were frozen after 48 hours by adding 100μL of well mixed culture to 100μL 30% glycerol in 96 well plates. Boxes containing heterozygotes grown in YPD+2μM nystatin were frozen after 72 hours while boxes containing heterozygotes grown in YPD+4μM nystatin were frozen after 96 hours.

6.6.3 *Sanger sequencing*

A small amount of freezer stock from 78 heterozygous replicates that showed growth was inoculated in 20mL of YPD and grown overnight for Sanger sequencing. Genomic DNA extractions were performed using standard protocols (Sambrook and Russell, 2001). We PCR amplified $\sim$1000bp around the mutation locus of interest and purified the product before Sanger sequencing in the forward direction. Ensembl and http://www.yeastgenome.org were used to obtain genome sequences, and the Clustalw multiple alignment tool from EBI was used for alignment (with default parameters, http://www.ebi.ac.uk/Tools/msa

/clustalw2/). Sequence Scanner Software v1.0 was used to look at the resulting chromatograms. All mutations and sequences were confirmed by alignment and visual inspection of the chromatograms.

6.6.4 *Illumina sequencing Aa-grow lines*

To assess the nature of LOH, genomic DNA from six lines that had lost heterozygosity but showed low G1 mean values (measured from flow cytometery) were resequenced in 100bp paired-end fragments using Illumina's HighSeq 2000 (BMN3-188, BMN9-12, BMN9-130, BMN13-54, BMN23-144, and BMN28-14). Library preps followed standard Illumina protocols (2011 Illumina, Inc., all rights reserved), with each line uniquely indexed and run together on a lane. The genomic sequence data were processed using Illumina's CASAVA-1.8.0. The CASAVA program configureBclToFastq.pl was used to convert to fastq and to separate the sequences by index (allowing one mismatched basepair per index).
configureAlignment.pl was then used to align each sequence to the yeast reference genome (scergenome.fasta downloaded from the Saccharomyces Genome Database,
http://downloads .yeastgenome.org/genomerelease/r64/). Finally, configureBuild.pl was used to call SNPs and indels and to obtain coverage data (see additional details in Chapter 4). Excluding mitochondrial DNA, the average depth of coverage was 60-fold (ranging from 16.5 to 384.8). Average coverage per chromosome is plotted in Figure E.3, and coverage in 1000 basepair windows across the genome is plotted in Figure E.4. In each case, inferences from Sanger sequencing about changes in the ergosterol genes were confirmed by CASAVA's SNP and indel calls.

Chapter 7

Discussion

"The only constant is change" - Heraclitus

7.1 Thesis Summary

Evolution enables species to adapt to a changing environment. Though natural selection operates at the phenotypic level, it is genomic variants that change over time in populations. However, the phenotypic effects of alleles are not constant, and may depend heavily on both the environmental and genomic background. In this thesis, I have attempted to explore both the range of mutations that might be available to populations, as well as how their effects may change in different genomic backgrounds and environments. I have focused my work on beneficial mutations, as these mutations, though paramount for evolution, have been less well characterized than deleterious mutations (likely because beneficial mutations are much rarer and so harder to isolate than deleterious mutations, as discussed in Chapter 6).

If we are ultimately interested in knowing how a given mutation enhances (or decreases) the fitness of the genome in which it appears, we must know whether more descendants carry that variant relative to other genotypes in the population. In Chapter 2, I describe a set of experiments designed to look for differences in fitness between haploid and diploid genotypes isolated from the same population at various time points. With prior knowledge of the outcome of this competition (diploids repeatedly overtook haploid populations, Gerstein *et al.* 2006), we thought it would be relatively straightforward to determine what character or characters allowed diploid mutants to outcompete haploids. Haploid and diploids colonies were isolated at multiple time points throughout the initial ~1800 generation experiment, utilizing one of the key benefits of microbial evolution experiments –

the ability to maintain and take advantage of the fossil record. We sampled colonies every $\sim$ 100 generations with additional colonies isolated at timepoints where polymorphism for ploidy was prevalent. We compared colonies of different ploidy using multiple fitness assays, including growth phase proxies and competitive fitness comparisons. Different fitness assays yielded inconsistent results, and surprisingly we did not find any fitness measures for which diploids repeatedly showed a clear advantage relative to haploids (Gerstein and Otto, 2011). We are left to conclude that the dynamics of diploid takeover were not due to a single character that differentiates all diploid colonies from all haploid colonies, and predict that frequency-dependent interactions may have contributed (discussed in greater detail below).

The work presented in Chapter 3 asked whether haploids or diploids adapted faster across seven different environments over $\sim$ 200 generations of batch culture evolution, a short timeframe where few mutations are expected to reach high frequency in any population. We used a competitive fitness assay to show that haploids adapted faster than diploids in seven different environments. We found that the environment significantly influenced the rate of adaptation, with tremendous variation among environments and between ploidy levels (Gerstein *et al.*, 2011). Although haploids adapted faster than diploids in all environments, this result was only significant in three environments, and we observed a lot of variation between replicate lines evolved in the same environment. The results of this experiment yielded perhaps more questions than answers, as many factors could have influenced the differences in rate we observed. In Gerstein *et al.* (2011) we present estimates of the dominance of the first selected mutations, finding our results consistent with the availability of dominant mutations in only four of seven environments. However, as discussed below, the assumptions we made may not always hold, and other explanations for the tendency of haploids to adapt faster, such as differences in mutation rate or mutation effect sizes, could also explain our results.

Chapter 4 describes the isolation and characterization of single beneficial mutations that confer resistance to 35 haploid lines in a single environment (the fungicide nystatin). We determined the genetic basis of adaptation for each line using whole genome resequencing. We found that the genomic scope for these first mutations involved in adaptation was narrow, as all lines carried one of 20 unique mutations in four different genes that act late in the ergosterol biosysnthesis pathway (*ERG7*, *ERG6*, *ERG5* and *ERG3*, Table 4.1). Different mutations in the same gene were found to share a similar tolerance to nystatin, regardless of mutation class (i.e., nonsynonymous SNP, small indel, large deletion). Very few secondary (non-ergosterol) mutations were identified, and our results are consistent with the ergosterol mutations largely (if not entirely) contributing to fitness improvements in nystatin. We found that performance in other stressful environments was variable among mutations in different genes, and although mutations in different genes sometimes behaved

similarly, the specific gene-environment interactions observed were not predictable.

The work presented in Chapter 5 and Chapter 6 of this dissertation sought to explore the effect of ploidy background on mutational parameters. In Chapter 5 I compared the effect of the nystatin adaptive mutations in haploids and homozygous diploids. Homozygous diploids, like haploids, lack a wildtype copy of the mutant allele, and it has previously been assumed in theoretical studies that the effect size of mutations should be the same in these two backgrounds. We compared mutations in these genomic backgrounds using parameters extracted from the nystatin dose-response relationship and growth rate assays. For these mutations and for multiple fitness parameters, we found that mutations in a haploid background had significantly larger effect sizes than in homozygous diploids. The precise relationship between mutations in different ploidy backgrounds, however, depended on both the gene carrying a mutation and the level of nystatin in the medium.

Novel mutations arise in a heterozygous state in diploids (i.e., a novel mutation appears in one gene copy while the second copy remains wild type), yet much remains unknown about the behaviour of adaptive mutations in heterozygous form. We thus constructed lines heterozygous for each of the twenty unique nystatin mutations. The initial goal of Chapter 6 was to measure the dominance of each mutation, i.e., whether heterozygotes behave more like wildtype or homozygous mutant diploids. In an unstressful environment we found considerable variation among mutations, ranging from seemingly dominant (the mutation in *ERG5*) to perfectly recessive (many of the other mutations). When we examined the mutations in heterozygous form under stressful conditions, we repeatedly found extremely inconsistent growth among replicates, even within a day. Through targeted Sanger sequencing we uncovered the source of this inconsistency: rapid and repeated loss of heterozygosity (LOH). This result was robust to the initial mutation and the level of stressor in the environment. This indicates that these mutations do not confer resistance to heterozygotes (i.e., they are recessive) under the levels of nystatin tested (from YPD+1μM nystatin to YPD+4μM nystatin), as heterozygous lines (that do not lose heterozygosity) are no more tolerant to nystatin than the wildtype, according to the fitness assays used.

The first two data chapters of this thesis described interesting patterns, yet without knowledge of the genetic basis of the mutations that were enabling adaptation it was difficult to make straightforward conclusions. The results of these earlier chapters thus motivated the work on single adaptive mutations conducted in the later chapters. This Ph.D. was undertaken at a fortuitous point in time. Midway through my dissertation, whole-genome resequencing (WGS) became both feasible and (somewhat) affordable for work in *S. cerevisiae*. Were it not for this stroke of luck, much of what was accomplished in Chapter 4 and 6 would not have been possible. In the last part of this Chapter I will discuss many of the research themes presented in this thesis, in the context of the results I have presented. Wherever possible (particularly for the topics touched on in the earlier chapters), I will up-

date the conclusions we made. Where it helps to draw a fuller picture, I also discuss results from other experiments I undertook during my Ph.D., but which were not included in this dissertation. The limitations of the dissertation research and a discussion of future research directions are presented throughout.

7.2 Why do Diploids Overtake Haploids? Revisiting Chapter Two

At the end of the experiments conducted in Chapter Two we were left with the result that the fitness assays we conducted did not show evidence that diploids were generally superior over haploids. We thus concluded that although ploidy appeared to be a deterministic factor (in that diploid mutants repeatedly outcompeted the haploids populations within which they arose, Gerstein *et al.* 2006), the precise dynamics of the route to which diploid colonies outcompeted haploid colonies was stochastic. As demonstrated in Chapter 2, without being able to precisely track a single genotype over long periods of time, it remains difficult to know exactly which fitness components (if any) best capture the complicated process of genotypic sorting over evolutionary time. As the genotypes within populations are themselves dynamic in both types and frequency, the genotype that wins may vary dependent on interactions with the other individuals present. This remains a major limitation of this study and other evolution experiments that seek to definitively declare what genotype is most fit within a long term experiment.

We have since conducted a series of direct competition experiments starting with different frequencies of haploids and diploids isolated from the same timepoints we focused on in Chapter 2 (haploid populations isolated from Gen1023 and Gen1302 and diploid populations from Gen1023, Gen1302, Gen1488 and Gen1767; note that diploidy had swept by Gen1488). As described in Chapter 2, we isolated and determined the ploidy of 24 colonies at each of these timepoints, thus ploidy populations were constructed for each time point as described in Chapter Two. These competition experiments showed repeated evidence for negative frequency dependence between ploidy populations (Figure 7.1), a further indication that complicated dynamics underlie the eventual diploid takeover. Similar dynamics are found when we compete single haploid or diploid genotypes isolated from Gen1302 (the last timepoint where polymorphism for ploidy was observed, data not shown). Whether or not ploidy itself enables these dynamics is unknown, it could be that ploidy is simply a marker that allows us to differentiate between two genotypes that exhibit frequency dependence.

It is surprising that diploids isolated from later time points (i.e., the populations from Gen1488 and Gen1767) were not able to overtake haploids, as these were populations of diploids isolated after diploidy had swept in the initial experiment. This is somewhat in

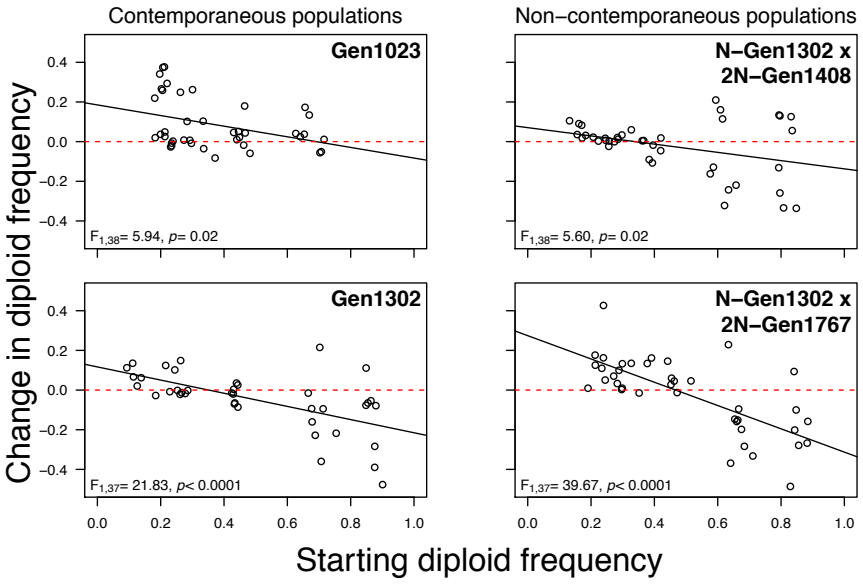


Figure 7.1: Negative frequency dependent selection is prevalent between ploidy populations isolated at different time points.

contrast with Figure 2.8, which shows the result of a competition between haploid and diploids populations from the same time points. As shown in Figure 7.1, although negative frequency dependence was found, for competitions that were initiated at or above diploid frequencies of $\sim 50\%$ we observed that different replicates seem to fall either above or below the predicted line. It may thus be that diploids by chance increased in frequency in the two replicates we evolved in Figure 2.8 or that there were microenvironmental differences between these experiments. The direct source of this variability among diploid lines remains unknown.

In an attempt to gain further insight into the potential mutational variants that might enable diploids to take over when they did, we whole genome re-sequenced the five haploid and five diploid colonies isolated after 1302 generations of evolution that we characterized in Chapter 2. This sequencing revealed a partial aneuploidy of chromosome five in the diploid colonies; two colonies were homozygous, two colonies were heterozygous, and one colony did not carry the aneuploidy. It would be fruitful in future experiments to determine the frequency of the aneuploidy at earlier time points (to determine when and where it arose, i.e., at what time point and in what ploidy background) and later time points (to determine if it swept through the diploid population). It is possible that this variant could be responsible for some of the complicated dynamics we observed. For example, it may be that diploid lines that carry the aneuploidy are able to outcompete haploids (explaining how diploids take over), yet are themselves outcompeted by diploids that do not carry

the aneuploidy (explaining why diploids from the final time point still display negative frequency dependence). It would also be interesting to cross the aneuploidy into the haploid background. As discussed in greater detail below, it may be that large chromosomal changes arise more often or are more beneficial in diploid backgrounds than in haploid backgrounds, and this could be a useful mutation with which to test this hypothesis.

7.3 The Rate of Adaptation is Influenced by Many Factors: Revisiting the Conclusions of Chapter Three

The work presented in Chapter 3 found that both ploidy and environment significantly influenced the rate of adaptation. Many factors directly affect the rate of adaptation, and could explain the main results we found and presented in this chapter, i.e., that haploids evolved somewhat to significantly faster than diploids in all environments, yet there was tremendous variation for the rate of adaptation between and among ploidy populations among environments. At the time this chapter was written and published (Gerstein *et al.*, 2011), we made the assumption that both mutation rate and mutation effect sizes are very similar between different ploidy populations to allow us to predict the dominance coefficients of the first acquired mutations in these lines. Our results were consistent with dominant to semi-dominant available mutations in only four of seven environments.

The results we later obtained in Chapter 5 suggest that the assumptions we made may not have been valid, at least for nystatin beneficial mutations, as we found beneficial mutations had larger effects in haploids than in diploid homozygotes. If mutations generally have a larger effect size in haploids, our estimates of dominance coefficients in Chapter 3 would have been underestimated. It may not, however, be a general result across environments that mutations have a larger effect size in haploids, nor is it necessarily the case that the mutations selected in the diploid lines are the same mutations that were selected in the haploid lines. Furthermore, it may or may not be the case that different mutations selected in different ploidy backgrounds have the same average effect size. Thus, without knowledge of the specific mutations that underlie adaptation, it is difficult to draw conclusions about the factors such as dominance that might have driven the main result of Chapter 3.

A second potentially confounding factor is the possibility of LOH within the diploid lines described in Chapter 3. As suggested in Chapter 6, if LOH frequently occurs then we may have overestimated the dominance coefficient of mutations (because a recessive mutation that has become homozygous will behave like a dominant beneficial mutation). It might be the case that LOH is induced at different rates under different stressors (as in Forche *et al.* 2011), and thus some of what we attributed to differences among environments

in dominance might be differences in the rate of LOH. It is possible that the mutations acquired in diploids in all environments were recessive, yet LOH occurred more frequently in YPD+KOH, YPD+Nystatin, YPD+NaCl and YPD+Caffeine (the four environments that were consistent with dominant mutations). Again, without knowing the zygosity of diploid mutations at the end time point of our experiment, it is difficult to tease apart the contributions of different mutation rates between haploids and diploids, different mutation effect sizes, the dominance of selected mutations, and the rate of LOH.

7.4 The Genetic Scope for Adaptation

In Chapter 4 we found that the genomic scope for the first mutations acquired that confer tolerance to nystatin was quite narrow, as we found mutations were present in only four different genes that act close together in the ergosterol biosynthesis pathway. The majority of mutations were non-synonymous SNPs (16/20 unique mutations), though we also identified indels and a 29bp duplication. The generality of these results remains unknown, as few experiments (particularly in eukaryotes) have characterized the first mutations acquired in response to a novel stressor. The scope and nature of beneficial mutations almost certainly depends on the environment, but will it commonly be the case that few genes are responsible and that most changes involve small-scale SNPs? Answers to these questions will only come from gathering data about adaptive mutations in response to different environmental challenges.

To address that aim, I recently conducted a parallel series of experiments in the same haploid strain background to obtain mutations that are adaptive in copper stress. Copper represents a considerably different challenge than does nystatin to yeast cells. Nystatin, a fungicide, acts by binding to ergosterol in the yeast cell membrane (Woods, 1971). We suspect that the mutations we acquired in nystatin are largely loss-of-function mutations that block the production of ergosterol and lead to the substitution of other sterols that are less affected by nystatin in the membrane (as has previously been found, Woods 1971; Hapala *et al.* 2005). We found that our lines had similar sterol profiles (Figure 4.2) and tolerance phenotypes (Figure 4.3) as lines that carry whole gene deletions for the appropriate genes (the one mutation in *ERG7* could not be assessed, as *ERG7* is an essential gene that is inviable as null). Copper, however, is an essential nutrient, yet it is highly toxic at high concentration (Peña *et al.*, 1999). Thus, unlike nystatin, cells cannot block copper uptake entirely.

We utilized the same mutation acquisition protocol (Chapter 4.3.1) and sequencing pipeline (Chapter 4.3.2) to acquire mutations in YPD+11mM $CuSO_4$. We found that the genetic basis of beneficial mutations in high copper was much broader than those acquired

in nystatin. A parallel genetic basis of adaptive targets was again found, with multiple lines each carrying (different) mutations in *VTC4* (10 lines), *VTC1* (3 lines), *PMA1* (5 lines) and *MAM3* (3 lines). We identified SNPs in 23 other genes, however, and these mutations seem to have a broad cellular basis (unlike our nystatin mutations, they act in many different pathways, though they primarily localize to the vacuole, endoplasmic reticulum, and mitochondria). We also identified 12 SNPs in non-coding regions of the genome. The majority of lines carried multiple mutations (between 1 and 10), and experiments are currently underway to isolate the fitness contributions of the different mutations in each line.

Another sharp contrast to the nystatin lines was found by looking for changes in copy number. We found widespread increases in copy number among copper adapted lines, both whole chromosomal aneuploidy and single gene duplication (by contrast, only a single nystatin line was found to contain chromosomal aneuploidy). Aneuploidy was identified in 12 different copper lines, each with an extra copy of either chromosome II (8 lines) or chromosome VIII (5 lines; one line had an extra copy of both). Five of these lines also had increased copies of one or two other chromosomes. Interestingly, 27 of 33 lines have 2-6 more copies of the *CUP1* locus than the wildtype; as *CUP1* is located on chromosome VIII, all but 2 lines thus showed an increased copy number of the *CUP1* locus (one line had both chromosome VIII aneuploidy and extra copies of *CUP1*). *CUP1* is known to be the major metallothionine activated under copper stress in yeast, and is the primary protein responsible for removing copper from the yeast cell (Karin *et al.*, 1984).

The cellular basis of copper tolerance likely explains why the copper mutation lines exhibited only slight increases in breadth of tolerance to varying concentrations of copper (Figure 7.2), compared to the large increase in tolerance to nystatin in our nystatin beneficial mutations (Figure 4.3). We suspect that the widespread changes in copy number we observed cause differences in gene expression of *CUP1*, and future experiments will characterize whether this is in fact the case. If possible, we will also characterize the molecular basis of other mutations identified in our copper screen. I believe this to be a worthwhile pursuit as few microbial evolution studies have fully linked beneficial mutations to their mode of action. I suspect we might acquire considerable insight into the evolutionary process (and its potential constraints) if we better understand the cellular and molecular basis by which beneficial mutations can act (this is not the first time this sentiment has been expressed, see for example Dalziel *et al.* 2009).

Combined, the results presented in Chapter 4 and our unpublished work on mutations that confer tolerance to copper demonstrate that the genomic scope for mutations likely depends heavily on the environment. We have not yet verified which of the mutations in our copper screen are adaptive, so the categorization of mutations may not be exact. Nevertheless, we found considerable variation in the type of mutations that arose between our two screens and among other studies that have similarly examined mutations that arose

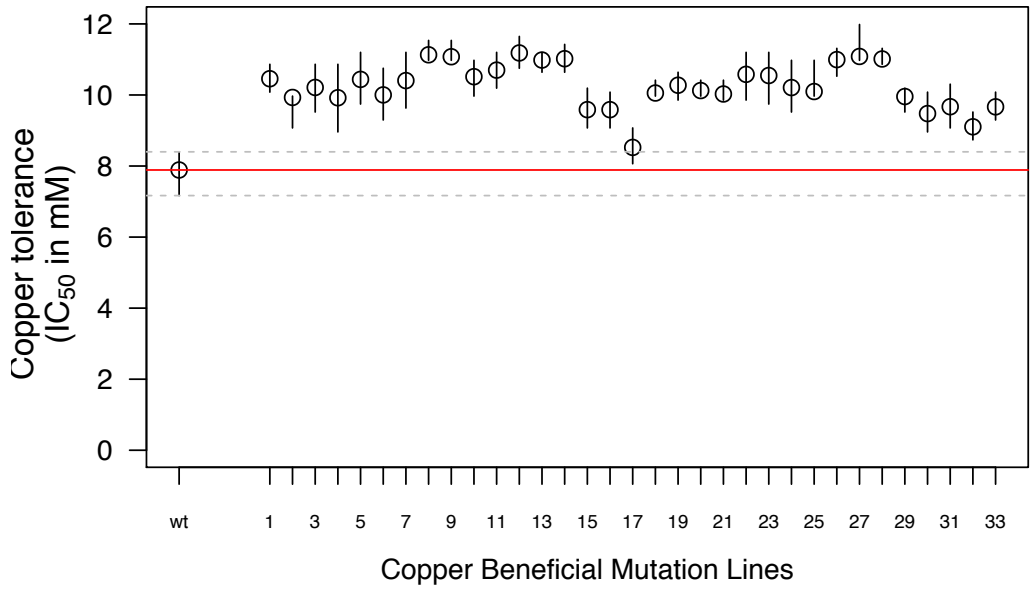


Figure 7.2: Copper adaptation lines exhibit only a slight increase in breadth of copper tolerance relative to the wildtype. Tolerance was measured as IC_{50}, as described in Chapter Four, section 4.3.4. Notice that the maximal increase in IC50 was always less than 1.5-fold for copper, but was up to ~80-fold in some nystatin lines (Figure 4.3).

within evolved microbial populations (Table 7.1, note I have not included studies from phage or viruses). The variation seen in Table 7.1 likely also reflects some species-specific differences. For example, IS-elements make up a considerable portion of changes in *E. coli* lines, whereas similar transposon-mediated changes are less often seen in other species. Additional structural mutations may have been present yet undetected in evolved lines, as bioinformatic tools poorly detect chromosomal rearrangements and inversions because they do not map well to reference genomes. Similarly, we were only able to identify the tandem duplications in *CUP1* in our copper screen by specifically looking for them (using the *in silico* qPCR technique described Chapter Six, section 6.3.4); without specifically looking for copy number changes in this exact gene, these duplications likely would have gone undetected. It is undoubtably the case that methods will be developed and available to the community in the next few years that will better enable researchers to search for these types of changes.

It is also possible that there are differences between the types of mutations that are selected at different stages of the evolutionary process, as the order that mutations arise has previously been shown to be important (Weinreich *et al.*, 2005). As discussed in Chapter 4, the majority of experimental evolution studies have identified mutations that arose within hundreds or thousands of generations. As shown in Table 7.1, a number of studies have, however, used fluctuation assays to isolate the first mutations that are acquired in response

Table 7.1: Classes of unique adaptive mutations acquired during experimental evolution studies identified by targeted sequencing (TS) or whole genome resequencing (WGS).

Species	Environment	Type of experiment	Sequencing	Number of mutations	SNPS/Indels/ Copy number/Other (%)	Authors
E. coli	glycerol	$batch_{44 days}$, 15 lines	WGS	13	62/23/15/0	Herring *et al.* 2006
E. coli	glucose minimal	$batch_{20000 gen}$, 5 lines[a]	WGS	55	27/18/0/35[b]	Barrick *et al.* 2009
E. coli	glucose minimal	$batch_{50 days}$, 9 lines	WGS	28	46/39/14/0	Charusanti *et al.* 2010
E. coli	temperature increase	$batch_{523 days}$, 15 lines	WGS	6	71/29/0/0	Kishimoto *et al.* 2010
E. coli	L-1,2-Propanediol	$batch_{700 gen}$, 1 line	WGS	13	33/33/0/33[c]	Lee and Palsson 2010
E. coli	isobutanol	$batch_{500 gen}$, 6 lines	WGS	105	79/16/0/5 [d]	Minty *et al.* 2011
E. coli	high temperature	$batch_{2000 gen}$, 115 lines	WGS	1331	63/32/2/5[e]	Tenaillon *et al.* 2012
E. coli	antibiotics[f]	$morbidostat_{20 days}$, 5 lines	WGS	40	78/0/10/ 13[g]	Toprak *et al.* 2011
P. aeruginosa	rifampicin	fluctuation, 80 lines	TS	15	93/17/0	MacLean and Buckling 2009
P. aeruginosa	quinolone	fluctuation, 18 lines	TS	9	89/11/0/0	Bataillon *et al.* 2011
P. aeruginosa	CF drugs[g]	$batch_{50 gen}$, 48 lines	WGS	63	62/37/0/2[h]	Wong *et al.*, in review
S. cerevisiae	sulfate limitation	$chemostat_{188 gen}$, 1 line	WGS	6	67/0/17/17[h]	Araya *et al.* 2010
S. cerevisiae	nystatin	fluctuation, 35 lines	WGS	20	80/15/5/0	Chapter Four
S. cerevisiae	copper	fluctuation, 33 lines	WGS	100	41/3/45/11[h]	Gerstein & Otto, unpublished

* : Copy number includes both duplications and chromosomal aneuploidy
[a] : clones isolated from 2K, 5K, 10K, 15K, 20K generations from the same population
[b] : 8 mutations not in coding regions, 10 IS-elements, 1 inversion
[c] : 1 IS-element, 1 change in 3 consecutive bp
[d] : 3 IS-elements, a 10kb deletion and a 1.7kb deletion
[e] : IS-elements
[f] : one of chloramphenicol, doxycycline, and trimethoprim
[g] : various combinations of synthetic CF sputum medium, mucin, and ciprofloxacin
[h] : mutations not in a coding regions

to a stressful environment. These assays are generally initiated by inoculating replicate cultures in a non-selective medium with a small number of cells that are allowed to grow up, and then plated onto selective medium to select for mutants that confer resistance to the selective environment (Luria and Delbrück, 1943). Our own experiments were a variation of the traditional fluctuation assays, in that we inoculated a small number of cells into stressful medium where they could only grow rapidly by acquiring a novel mutation.

The five studies listed in Table 7.1 that acquired mutations from fluctuation assays tended to acquire a higher proportion of SNPs than the studies that evolved lines for much longer. The exception to this trend, our study in copper, is heavily influenced by 27 cases of a *CUP1* duplication, which seems to be directly selected by the specific environmental pressure. It may be that insertions, deletions and copy number changes (which are more frequent in the longer-term studies) are not generally the first mutations to arise when populations are subjected to a novel stressor. An experiment could be designed to test this. The critical component is the ability to isolate mutations as they arise within populations over extended periods of time. A method developed by Kao and Sherlock (2008) could facilitate such efforts. Their method utilizes strains of yeast that have the same genotype, yet are constructed to contain different fluorescent markers. By tracking the marker frequencies over time, it is possible to determine when new mutations have arisen (i.e., when the frequency of one marker increases relative to the others). In this way, if populations were

examined daily for the first 1-14 days after experiments were initiated, one could potentially pick out the earliest mutations that arose in replicate lines, while allowing the experiment to run for a long period of time. The genetic basis of the first mutations could then be compared to the pool of mutations that sweep at later time points. In a similar vein, it would be interesting to evolve the nystatin lines further to explore the second-step mutations. Will the second-step mutations also be concentrated among the same four genes (especially, within *ERG6* and *ERG3*)? Is the scope for secondary mutations affected by the first mutations that arose within the lines? How much are populations constrained by the first mutations that appear in the genome? Are the second step mutations largely compensatory, with respect to the pleiotropic effects of the first mutation?

The level of stressor under which mutations were acquired in might also be an important determinant on the types of mutations that are selected. Although a major goal of the work presented in Chapter 4 was to evaluate the genomic scope for mutations, we were limited in this endeavour by the use of fluctuation assays that acquire mutations in a level of stressor at which the wildtype is unable to grow, and are thus biased towards major effect rescue mutations. Yet it is not necessarily the case that the mutations that enable adaptation to higher levels of a stressor will necessarily be the same class of mutations or in the same genes as those that would be acquired at lower levels of stress. It may well be the case that smaller effect size mutations were available that could improve tolerance to slight increases in nystatin, yet these were not selected in our screen. To my knowledge, few studies have directly asked whether different mutations are acquired at different levels of the same stressor, yet experiments on the dominance of adapted lines by Anderson *et al.* (2004) indicate this may well be the case, at least for some stressors. If evolutionary biologists working with microbes hope to shed light on evolutionary processes in 'the real world' (and I believe they do), we should examine the effect that different levels of stressors have on our results. Many of our experiments (including all of my own) subject ancestral culture to high levels of stressor without a 'ramp-up' period. More realistic experiments may be those that gradually increase levels of stress over time, e.g., as in Anderson *et al.* (2003) and Bell and Gonzalez (2011).

A follow-up experiment with evolved lines from our lab could also potentially shed light on this question. In the experiment described in Chapter 3, haploid lines were evolved in 0.6μM nystatin. We could thus do targeted resequencing of the four genes we identified in our screen at 4μM nystatin to look for mutations in these same genes in culture from these lines that was frozen down after $\sim$ 50, 100, 150 and 200 generations of evolution at this much lower level of nystatin. It would be extremely interesting to determine the relationship between stage of evolutionary process, level of stressor, type of mutation, and mutational mode of action (i.e., loss of function versus gain of function), and whether there are interactions between them. This knowledge could inform our understanding of

the abilities and constraints of populations adapting to stressful environments. It may or may not be that different mutations arise in populations at either different stages of evolution or due to different levels of environments stressor, and these mutations may constrain (or enhance) the number of evolutionary pathways available to future populations. As sequencing technology gets cheaper (and the bioinformatic tools to analyze the data get easier), it seems likely that the scope of possible experiments will rise exponentially in the upcoming years.

7.5 Ploidy Background Influences the Effect Size of Beneficial Mutations

The majority of microbial evolution studies have been done on haploid species or populations. Yet haploids represent just one group of life on the planet, and it may well be that the conclusions about the properties of mutations that we make from studying haploids can not be extrapolated to diploids. The ability to manipulate the ploidy of experimental lines is not possible in many of the focal species used in microbial evolution studies, yet is possible with *S. cerevisiae*. I thus manipulated the ploidy background of the twenty unique nystatin mutations we acquired in Chapter 4 to specifically determine how ploidy affects mutation effect size. In Chapter 5 we demonstrated that, for multiple aspects of fitness, these mutations had a significantly larger effect size in haploids compared to their effects in homozygous diploids.

If this result is repeatable across many environments and in multiple species, it indicates a potentially major benefit of haploidy. However, it remains unknown whether the same mutations that appear in haploid genomes are also those utilized by diploids. It may be that diploids have access to a different set of mutations than haploids, and thus, on average, the net rate of adaptation in different ploidy backgrounds might be similar. For example, as discussed in Chapter 5, the data that exist suggest that large chromosomal changes may frequently be beneficial only in a diploid background (Thompson *et al.*, 2006; Gresham *et al.*, 2008). A confounding factor in the examination of mutations acquired in diploids using the techniques described above (fluctuation assays, long-term evolution experiments, competition between fluorescently labelled strains) is that primarily dominant beneficial mutations are expected to be selected *de novo* in diploid populations (Haldane, 1927). As discussed in Chapter 6, we do not have a good sense of the dominance properties of adaptive mutations. Mutations acquired *de novo* in diploids using common techniques might thus be biased against recessive mutations, which might nevertheless be important to the evolutionary process if they are frequently maintained as standing genetic variation. Whether the mutations important to adaptation typically arise *de novo* or as standing genetic variation

also remains an outstanding question (Hermisson and Pennings, 2005). Acquiring mutations in a haploid background (as we have done) circumvents this problem, yet remains a different limitation if diploid-specific beneficial mutations exist and are not acquired.

Future experiments will be able to shed light on whether the same mutations arise in haploid and diploid populations. As discussed above, WGS could be done on the haploid and diploid lines we evolved in Chapter 3. Since we do not expect very many mutations to be segregating at high frequency in these lines (Figure B.2), this sequencing could allow us to determine whether the first mutations that are selected are different in haploid and diploid lines. If different classes of mutations are indeed found, experiments similar to those in Chapter 5 would tell us whether these mutations have different effect sizes in haploids and diploids. Similar experiments could be done over a longer timescale to see whether the same pattern holds for the first-step mutations as later-step mutations.

7.6 What is the Distribution of Dominance Coefficients for Beneficial Mutations?

The effects of deleterious mutations in heterozygous form has long been studied by evolutionary biologists (e.g., Crow and Temin, 1964; Mukai *et al.*, 1972), with empirical evidence from many species indicating that deleterious mutations are, on average, recessive. Why this should be the case, however, led to a prominent debate between Fisher (1928) and Wright (1934). Kacser and Burns (1981) published a seminal paper on metabolic control theory (MCT) that demonstrated that wildtype alleles were dominant over deleterious alleles due to an innate property of the kinetics of metabolic systems. They demonstrated that intermediate changes in enzyme concentration due to heterozygous mutations at any step along the metabolic pathways were unlikely to have an effect on the output of the system (i.e., metabolic flux or other measurable phenotypes). One of the motivating questions of this dissertation was whether beneficial mutations would also be largely recessive.

In Chapter 6 we sought to assess the dominance of the nystatin adaptive mutations acquired in Chapter 4. We found considerable variation among mutations when we assayed them in an unstressful environment in which the haploid mutations were previously found to be deleterious (Chapter 4). Mutations spanned the range of possible dominance values, from seemingly dominant to perfectly recessive. The assay used to measure dominance in the unstressful environment was a growth rate assay that examined both maximal growth rate and the biomass production after 48 hours of growth. Somewhat surprisingly, we found a lot of variation between the two days this experiment was run. Though different replicates of the same line tended to behave very similarly within an experiment, results were often inconsistent between days (as evidenced by the large error bars in Table 6.1). A similar day

effect was found by (Agrawal and Whitlock, 2011), who re-examined competitive fitness data from the set of yeast deletion collection lines (Deutschbauer, 2005).

The source of this day-to-day variation remains unknown. Furthermore, it is not known how often such variation might be present, as many projects do not repeat their experiments on multiple days. Early in my Ph.D., I conducted a series of experiments with the aim to quantify day-to-day variation in growth of haploid and diploid wildtype lines (the ancestral colonies of the experiments described in Chapter 3). These experiments also examined grown in YPD, the same unstressful environment. As shown in Figure 7.3, considerable variation in maximal growth rate exists among wildtype cultures grown in different batches of medium, even if measured in the bioscreen on the same day. Medium batch was a highly significant predictor of growth rate ($F_5 = 6.40$, $p < 0.0001$), while neither ploidy, nor their interaction were significant (ploidy: $F_2 = 0.15$, $p = 0.86$; interaction: $F_8 = 0.78$, $p = 0.62$). Throughout this thesis, I thus never compared bioscreen results started on different days or using different batches of medium.

Despite the increased noise caused by day-to-day variation in our replicate fitness measures, we were able to document the biologically important result, that there exists differences in the dominance of nystatin mutations for growth in YPD. This result should be examined further. Future experiments will use a competitive fitness assay (as presented in Chapter 3) to directly compare the fitness of homozygous and heterozygous mutant diploids against wildtype diploids.

When we examined the mutations in heterozygous form under stressful conditions (which was the original goal) we found extremely inconsistent growth among all replicates, even those assayed within a single day. We also observed that many replicates did not grow at all; combined these dynamics appeared very similar to the pattern of stochastic growth used to acquire new mutations in Chapter 4. Targeted Sanger sequencing uncovered rapid and repeated loss of heterozygosity of the initial mutations. We thus conclude that these mutations are recessive under the levels of nystatin we tested; in no environment did we find that the heterozygous lines (that did not lose heterozygosity) were more tolerant than the wildtype to nystatin.

Other studies in yeast have also found evidence for rapid loss of heterozygosity of adaptive mutations (Gresham *et al.*, 2008; Anderson *et al.*, 2004). It may thus be difficult to quantify the fitness of heterozygotes (and thus assess the dominance of mutations) without directly examining whether LOH is present in fitness assays. Dominance measures are particularly problematic in cases where heterozygous fitness is much lower than that of a homozygote, because selection will rapidly amplify lineages with the appropriate LOH. As mitotic recombination has previously been shown to be the primary mechanism enabling loss of heterozygosity in yeast (St Charles *et al.*, 2012; Hiraoka *et al.*, 2000), it might be

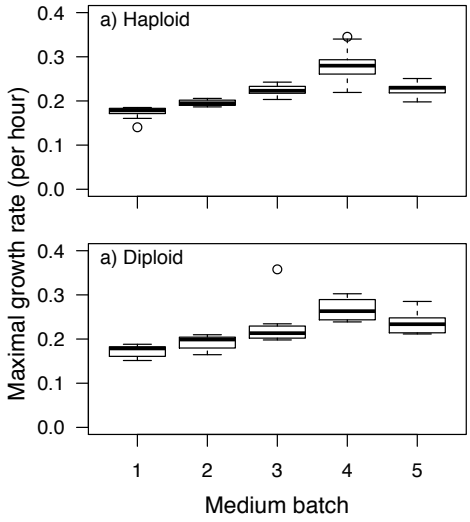


Figure 7.3: Growth rate is considerably affected by medium batch. Five batches of YPD were mixed and autoclaved on different days (except batches 3 & 4, which were mixed and autoclaved separately on the same day). Wildtype haploid and diploid culture (the same genotype used in Chapter 3) was struck from frozen culture onto YPD plates and allowed to grow for 72 hours. A single colony of each ploidy was then picked into a single 10mL test tube and grown overnight. 1:101 dilutions from these common tubes were then used to inoculate medium from each batch, and run in a single bioscreen. Show are standard boxplots, each based on 12 replicates.

possible to lessen the frequency of LOH if we conducted similar experiments on lines that carry knockouts of the primary genes required for recombination (e.g., *RAD52*, Freedman and Jinks-Robertson 2002). Alternatively, fitness assays involving fewer cells could be conducted to avoid the appearance of LOH cells (e.g., colony growth assays or FACs counts of liquid cultures starting with fewer cells).

An estimation of the dominance of beneficial mutations remains an important goal. If stable heterozygotes can be obtained, future experiments in yeast could aim to measure the dominance of beneficial mutations under different environmental conditions and under different levels of the same stressor. An important study by Anderson *et al.* (2004) demonstrated that the dominance of selected mutations may depend on the level of stressor. Some of my data also speak to this possibility. Haploids and homozygous diploids carrying mutations in *ERG3* and *ERG7* grew stochastically in YPD+8μM nystatin (Chapter 5) as well as YPD+16μM nystatin (data not shown). By examining the number of lines with mutations in *ERG3* and *ERG7* that showed stochastic growth in these two environments, I uncovered an interesting result. Not surprisingly, fewer mutations arose within lines grown in YPD+16μM (Figure 7.4). However, what I didn't expect was that the ratio of haploid to diploid lines that showed growth also changed between environments. In YPD+8μM nystatin, haploid lines grew more often than diploid lines, while in YPD+16μM nystatin there much less of a differences between the rate that lines of different ploidy acquired mutations. To test whether this trend was significant, I ran a generalized linear model with a binomial error distribution (and logistic link function). The presence of a mutation was affected by both the environment ($p = 0.032$) and ploidy ($p < 0.0001$), and the interaction between them was also significant ($p = 0.011$, Table 7.2). This result thus predicts that the mutations that enable growth at YPD+16μM nystatin are more often dominant than mutations that enable growth at YPD+8μM nystatin or that there is a broader class of mutations accessible to diploids at the higher concentration of nystatin. This work could be followed up by sequencing lines that grew at both environments, isolating the mutations within and measuring dominance as in Chapter 6. This would be another way to examine whether the same mutations generally arise within haploid and diploid lines and could also be used to test directly whether dominant mutations arise more often in diploid lines than in haploid lines (as predicted by Haldane's sieve).

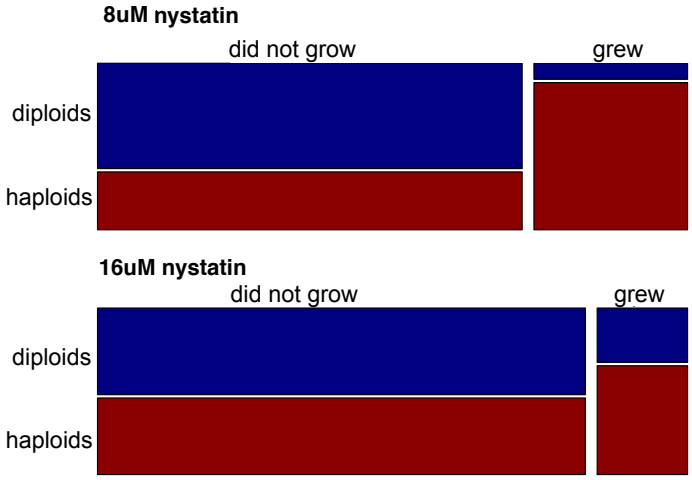


Figure 7.4: Haploid lines grew (and presumably acquired mutations) significantly faster than diploid lines in YPD+8μM nystatin (top mosaic plot), but not in YPD+16μM nystatin (bottom). This result potentially indicates the mutations acquired at YPD+16μM nystatin are more often dominant than those acquired at YPD+8μM nystatin. 96 replicates were assessed for each ploidy level in YPD+8μM nystatin while 48 replicates were assessed for each ploidy level in YPD+16μM nystatin.

Table 7.2: Environment, ploidy and their interaction significantly affect the likelihood of a secondary mutation in *ERG3* and *ERG7* haploids and diploids grown in YPD+8μM nystatin and YPD+16μM nystatin

	Df	Residual Deviance	p-value
null	310.04	287	
environment	305.49	286	0.033
ploidy	259.82	286	< 0.0001
interaction	253.41	284	0.011

7.7 Conclusions

The pursuit of scientific knowledge can be a frustrating enterprise. Though laws may exist in some scientific disciplines (e.g., physics), as a biologist it often seems that there are more exceptions than rules. The truth is, evolutionary dynamics are complicated, even for a unicellular organism adapting to life in a test tube. A large part of this thesis came about by embracing the unexpected (albeit, often reluctantly). Sometimes during the work done for this dissertation it seemed that each new result simply led to an exponentially larger number of new, unanswered questions. I was lucky to be able follow up on some of these unexpected results. Working with microorganisms allowed me to keep a fossil (i.e., freezer) record of the intermediate steps of evolution, an advantage not found in many biological systems. Furthermore, working in an organism with a small genome, it became possible to 'peer under the hood of evolution', to examine what had enabled change at the genomic level. Although this genomic knowledge will not always inform our unanswered questions (i.e., why does the phenotype differ between mutations in haploids and homozygous diploids, Chapter 5), in many cases it can (i.e., heterozygotes replicates showed stochastic growth because some replicates had rapidly lost heterozygosity, Chapter 6). Many of the experiments I have proposed in this last chapter describe future experiments that could use sequencing technology to potentially help explain the patterns we found (e.g., differences in the rate of adaptation between haploids and diploids, Chapter 3).

It seems that as evolutionary biologists, we ought to expect the unexpected. The first data chapter of this thesis (Chapter 2) documents a case where what we expect to be very different genotypes (haploids and diploids isolated from polymorphic populations within an ~1800 generation evolution experiment), had very little phenotypic or fitness differences. By contrast, however, when we examined the effects of single mutations in these same two backgrounds, we found that the set of adaptive mutations we examined had a much larger effect in haploids compared to diploids (Chapter 5). We also documented an example where very parallel genotypes (mutations in four genes that act very close together at the end of a pathway) exhibited either very similar or very different phenotypes, depending on the environment in question (Chapter 4). The work done for this thesis has sought to reinforce the notion that the relationship between genotype, phenotype, and fitness is complicated, and likely depends very much on both the genomic background and the environment.

Bibliography

Adams, J., and P. Hansche, 1974 Population studies in microorganisms I. Evolution of diploidy in *Saccharomyces cerevisiae*. Genetics **76**: 327–338.

Adams, J., C. Paquin, P. Oeller, and L. Lee, 1985 Physiological characterization of adaptive clones in evolving populations of the yeast, *Saccharomyces cerevisiae*. Genetics **110**: 173–185.

Agrawal, A. F., and M. C. Whitlock, 2011 Inferences about the distribution of dominance drawn from yeast gene knockout data. Genetics **187**: 553–566.

Aguilera, F., R. Peinado, C. Millán, and J. Ortega, 2006 Relationship between ethanol tolerance, H+-ATPase activity and the lipid composition of the plasma membrane in different wine yeast strains. Int. J. Food Microbiol. **110**: 34–42.

Amberg, D. C., D. J. Burke, and J. N. Strathern, 2005 *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*. CSHL Press, Long Island, New York.

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. Curr Biol **20**: 1383–1388.

Anderson, J. B., C. Sirjusingh, A. B. Parsons, C. Boone, C. Wickens, *et al.*, 2003 Mode of selection and experimental evolution of antifungal drug resistance in *Saccharomyces cerevisiae*. Genetics **163**: 1287–1298.

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

Applebee, M. K., A. R. Joyce, T. M. Conrad, D. W. Pettigrew, and B. O. Palsson, 2011 Functional and metabolic effects of adaptive glycerol kinase (glpk) mutants in *Escherichia coli*. J Biol Chem **286**: 23150–23159.

Araya, C. L., C. Payen, M. J. Dunham, and S. Fields, 2010 Whole-genome sequencing of a laboratory-evolved yeast strain. BMC Genomics **11**: 88.

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*. J Clin Microbiol **37**: 3332–3337.

Bard, M., 1972 Biochemical and genetic aspects of nystatin resistance in *Saccharomyces cerevisiae*. J Bacteriol **111**: 649–657.

Bard, M., J. L. Neuhauser, and N. D. Lees, 1980 Caffeine resistance of *Saccharomyces cerevisiae*. J Bacteriol **141**: 999–1002.

Barrett, R. D. H., L. K. M'gonigle, and S. P. Otto, 2006 The distribution of beneficial mutant effects under strong selection. Genetics **174**: 2071–2079.

Barrett, R. D. H., and D. Schulter, 2007 Adaptation from standing genetic variation. TREE **23**: 38–44.

Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, *et al.*, 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature **461**: 1243–1247.

Bataillon, T., T. Zhang, and R. Kassen, 2011 Cost of adaptation and fitness effects of beneficial mutations in *Pseudomonas fluorescens*. Genetics **189**: 939–949.

Bell, G., and S. Collins, 2008 Adaptation, extinction and global change. Evo. Appl. **1**: 3–16.

Bell, G., and A. Gonzalez, 2011 Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration. Science **332**: 1327–1330.

Bennett, A., and R. Lenski, 2007 An experimental test of evolutionary trade-offs during temperature adaptation. PNAS **104**: 8649–8654.

Bhiyan, M. S. A., Y. Ito, A. Nakamura, N. Tanaka, K. Fujita, *et al.*, 1999 Nystatin effects on vacuolar function in *Saccharomyces cerevisiae*. Bioscience **63**: 1075–1082.

Blasina, A., B. Price, G. Turenne, and C. McGowan, 1999 Caffeine inhibits the checkpoint kinase ATM. Curr Biol **9**: 1135–1138.

Blount, Z., C. Borland, and R. Lenski, 2008 Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. PNAS **105**: 7899–7906.

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

Bourguet, D., and M. Raymond, 1998 The molecular basis of dominance relationships: the case of some recent adaptive genes. J Evol Biol **11**: 103–122.

Burch, C., and L. Chao, 2000 Evolvability of an RNA virus is determined by its mutational neighbourhood. Nature **406**: 625–628.

Campos, P. R. A., P. S. C. A. Neto, V. M. D. Oliveira, and I. Gordo, 2008 Environmental heterogeneity enhances clonal interference. Evolution **62**: 1390–1399.

Campos, P. R. A., and L. M. Wahl, 2009 The effects of population bottlenecks on clonal interference, and the adaptation effective population size. Evolution **63**: 950–958.

Carrillo-Munoz, A. J., G. Giusiano, P. A. Ezkurra, and G. Quindós, 2006 Antifungal agents: mode of action in yeast cells. Rev Esp Quimioter **19**: 130–139.

Cavenee, W., T. Dryja, R. Phillips, W. Benedict, R. Bodbout, *et al.*, 1983 Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. Nature **305**: 771–775.

Charusanti, P., T. M. Conrad, E. M. Knight, K. Venkataraman, N. L. Fong, *et al.*, 2010 Genetic basis of growth adaptation of *Escherichia coli* after deletion of pgi, a major metabolic gene. PLoS Genetics **6**: e1001186.

Chavarrias, D., C. Lopez-Fanjul, and A. Garcia-Dorado, 2001 The rate of mutation and the homozygous and heterozygous mutational effects for competitive viability: A long-term experiment with *Drosophila melanogaster*. Genetics **158**: 681–693.

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 Res **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 Genet **5**: e1000652.

Christin, P.-A., D. M. Weinreich, and G. Besnard, 2010 Causes and evolutionary significance of genetic convergence. Trends Genet **26**: 400–405.

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 Biol **10**: R118.

Conrad, T. M., N. E. Lewis, and B. Palsson, 2011 Microbial laboratory evolution in the era of genome-scale science. Mol Sys Biol **7**: 1–11.

Cooper, T. F., D. E. Rozen, and R. E. Lenski, 2003 Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. PNAS **100**: 1072–1077.

Cowen, L., L. Kohn, and J. Anderson, 2001 Divergence in fitness and evolution of drug resistance in experimental populations of *Candida albicans*. J Bacteriol **183**: 2971–2978.

Crow, J., and R. Temin, 1964 Evidence for the partial dominance of recessive lethal genes in natural populations of Drosophila. Am Nat **98**: 21–33.

Cuevas, J. M., S. F. Elena, and A. Moya, 2002 Molecular basis of adaptive convergence in experimental populations of RNA viruses. Genetics **162**: 533–542.

Dalziel, A. C., S. M. Rogers, and P. M. Schulte, 2009 Linking genotypes to phenotypes and fitness: how mechanistic biology can inform molecular ecology. Molecular Ecology **18**: 4997–5017.

de Godoy, L., J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, *et al.*, 2008 Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature **455**: 1251–1254.

de Visser, J., and R. Lenski, 2002 Long-term experimental evolution in *Escherichia coli*. xi. rejection of non-transitive interactions as cause of declining rate of adaptation. BMC Evol Biol **2**: 19–27.

de Visser, J., and D. Rozen, 2006 Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. Genetics **172**: 2093–2100.

Desai, M., D. Fisher, and A. Murray, 2007 The speed of evolution and maintenance of variation in asexual populations. Curr Biol **17**: 385–394.

Deutschbauer, A. M., 2005 Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics **169**: 1915–1925.

Dickinson, W. J., 2008 Synergistic fitness interactions and a high frequency of beneficial changes among mutations accumulated under relaxed selection in *Saccharomyces cerevisiae*. Genetics **178**: 1571–1578.

Dujon, B., 2010 Yeast evolutionary genomics. Nature **11**: 512–524.

Dykhuizen, D., 1990 Experimental studies of natural-selection in bacteria. Annu Rev Eco Sys **21**: 373–398.

Dykhuizen, D. E., and D. L. Hartl, 1983 Functional effects of pgi allozymes in *Escherichia coli*. Genetics **105**: 1–18.

Ezov, T. K., E. Boger-Nadjar, Z. Frenkel, I. Katsperovski, S. Kemeny, *et al.*, 2006 Molecular-genetic biodiversity in a natural population of the yeast saccharomyces cerevisiae from "evolution canyon": Microsatellite polymorphism, ploidy and controversial sexual status. Genetics **174**: 1455–1468.

Fisher, R., 1928 The possible modification of the response of the wild type to recurrent mutations. Am Nat **62**: 115–116.

Fisher, R. A., 1930 *The genetical theory of natural selection*. Oxford: Oxford University Press.

FitzJohn, R. G., W. P. Maddison, and S. P. Otto, 2009 Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst Biol **58**: 595–611.

Forche, A., D. Abbey, T. Pisithkul, M. A. Weinzierl, T. Ringstrom, *et al.*, 2011 Stress alters rates and types of loss of heterozygosity in *Candida albicans*. mBio **2**: e00129–11.

Freedman, J. A., and S. Jinks-Robertson, 2002 Genetic requirements for spontaneous and transcription-stimulated mitotic recombination in *Saccharomyces cerevisiae.* Genetics **162**: 15–27.

Fry, J., 2004 On the rate and linearity of viability declines in Drosophila mutation-accumulation experiments: genomic mutation rates and synergistic epistasis revisited. Genetics **166**: 797–806.

Galitski, T., A. Saldanha, C. Styles, E. Lander, and G. Fink, 1999 Ploidy regulation of gene expression. Science **285**: 251–254.

Garcia-Dorado, A., and A. Caballero, 2000 On the average coefficient of dominance of deleterious spontaneous mutations. Genetics **155**: 1991–2001.

Gerrish, P., and R. Lenski, 1998 The fate of competing beneficial mutations in an asexual population. Genetica **102-3**: 127–144.

Gerstein, A. C., H.-J. E. Chun, A. Grant, and S. P. Otto, 2006 Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. PLoS Genet. **2**: e145.

Gerstein, A. C., L. A. Cleathero, M. A. Mandegar, and S. P. Otto, 2011 Haploids adapt faster than diploids across a range of environments. J Evol Biol **24**: 531–540.

Gerstein, A. C., R. M. McBride, and S. P. Otto, 2008 Ploidy reduction in *Saccharomyces cerevisiae*. Biol Lett **4**: 91–94.

Gerstein, A. C., and S. P. Otto, 2009 Ploidy and the causes of genomic evolution. J Hered **100**: 571–581.

Gerstein, A. C., and S. P. Otto, 2011 Cryptic fitness advantage: diploids invade haploid populations despite lacking any apparent advantage as measured by standard fitness assays. PLoS One **6**: e26599.

Ghannoum, M. A., and L. B. Rice, 1999 Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. Clin Microbiol Rev **12**: 501–517.

Giannattasio, S., N. Guaragnella, M. Cortereal, S. Passarella, and E. Marra, 2005 Acid stress adaptation protects from acetic acid-induced programmed cell death. Gene **354**: 93–98.

Giaver, G., A. M. Chu, L. Ni, C. Connelly, and L. R. *et al.*, 2002 Functional profiling of the *Saccharomyces cerevisiae* genome. Nature **418**: 387–391.

Gillespie, J., 1991 *The Causes of Molecular Evolution*. Oxford University Press, New York.

Gillespie, J. H., 1983 A simple stochastic gene substitution model. Theor Popul Biol **23**: 202–215.

Gillespie, J. H., 1984 Molecular evolution over the mutational landscape. Evolution **38**: 1116–1129.

Glazunov, A. V., A. V. Borelko, and A. K. Esser, 1989 Relative competitiveness of haploid and diploid yeast cells growing in a mixed population. Mikrobiologiia **58**: 769–777.

Goffeau, A., B. Barrell, H. Bussey, R. Davis, B. Dujon, *et al.*, 1996 Life with 6000 genes. Science **274**: 546, 563–567.

Gompel, N., and B. Prud'homme, 2009 The causes of repeated genetic evolution. Dev Biol **332**: 36–47.

Gould, S. J., and R. C. Lewontin, 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B **205**: 581–98.

Gregory, T. R., 2005 Synergy between sequence and size in large-scale genomics. Nat Rev Genet **6**: 699–708.

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 Genet **4**: e1000303.

Grunwald-raij, H., and P. Margalith, 1990 Ethanol fermentation by nystatin-resistant strains of Saccharomyces cerevisiae. Journal of Applied Microbiology **68**: 247–252.

Haldane, J. B. S., 1927 A mathematical theory of natural and artificial selection, Part V: Selection and mutation. P Camb Philos Soc **23**: 838–844.

Hall, D., and S. B. Joseph, 2010 A high frequency of beneficial mutations across multiple fitness components in *Saccharomyces cerevisiae*. Genetics **185**: 1397–1409.

Hannan, M. A., and A. Nasim, 1977 Caffeine enhancement of radiation killing in different strains of *Saccharomyces cerevisiae*. Mol. Gen. Genet. **158**: 111–6.

Hapala, I., V. Klobucnikova, K. Mazanova, and P. Kohut, 2005 Two mutants selectively resistant to polyenes reveal distinct mechanisms of antifungal activity by nystatin and amphotericin b. Biochem Soc Trans **33**: 1206–1209.

Harr, B., M. Kauer, and C. Schlötterer, 2002 Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. PNAS **99**: 12949–12954.

Hereford, J., 2009 A quantitative survey of local adaptation and fitness trade-offs. Am Nat **173**: 579–588.

Hermisson, J., and P. Pennings, 2005 Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. Genetics **169**: 2335.

Herring, C. D., A. Raghunathan, C. Honisch, T. Patel, M. K. Applebee, *et al.*, 2006 Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat Genet **38**: 1406–1412.

Hillenmeyer, M. E., E. Fung, J. Wildenhain, S. E. Pierce, S. Hoon, *et al.*, 2008 The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. Science **320**: 362–365.

Hiraoka, M., K. Watanabe, K. Umezu, and H. Maki, 2000 Spontaneous loss of heterozygosity in diploid *Saccharomyces cerevisiae cells*. Genetics **156**: 1531–1548.

Houle, D., K. Hughes, and S. Assimacopoulos, 1997 The effects of spontaneous mutation on quantitative traits. ii. Dominance of mutations with effects on life-history traits. Genet Res **70**: 27–34.

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–e00199–17.

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. Antimicrob Agents Ch **42**: 1160–1167.

Joseph, S. B., and D. Hall, 2004 Spontaneous mutations in diploid *Saccharomyces cerevisiae*: More beneficial than expected. Genetics **168**: 1817–1825.

Joyce, P., D. R. Rokyta, C. J. Beisel, and H. A. Orr, 2008 A general extreme value theory model for the adaptation of dna sequences under strong selection and weak mutation. Genetics **180**: 1627–1643.

Kacser, H., and J. Burns, 1981 The molecular basis of dominance. Genetics **97**: 639–666.

Kanafani, Z. A., and J. R. Perfect, 2008 Resistance to antifungal agents: mechanisms and clinical impact. Clin Infect Dis **46**: 120–128.

Kao, K. C., and G. Sherlock, 2008 Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. Nat Genet **40**: 1499–1504.

Kapteyn, J., B. ter Riet, E. Vink, S. Blad, H. D. Nobel, *et al.*, 2001 Low external pH induces HOG1-dependent changes in the organization of the *Saccharomyces cerevisiae* cell wall. Mol Microbiol **39**: 469–479.

Karin, M., R. Najarian, A. Haslinger, P. Valenzuela, J. Welch, *et al.*, 1984 Primary structure and transcription of an amplified genetic locus: the *CUP1* locus of yeast. PNAS **81**: 337–341.

Kassen, R., and T. Bataillon, 2006 Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat Genet **38**: 484–8.

Kim, D.-U., J. Hayles, D. Kim, V. Wood, H.-O. Park, *et al.*, 2010 Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. Nat Biotechnol **28**: 617–623.

Kimura, M., 1968 Evolutionary rate at the molecular level. Nature **217**: 624–626.

Kimura, M., and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. Genetics **49**: 725–38.

Kirkpatrick, M., and C. D. Jenkins, 1989 Genetic segregation and the maintenance of sexual reproduction. Nature **339**: 300–301.

Kishimoto, T., L. Iijima, M. Tatsumi, N. Ono, A. Oyake, *et al.*, 2010 Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution. PLoS Genet **6**: e1001164.

Kondrashov, A., and J. Crow, 1991 Haploidy or diploidy - which is better. Nature **351**: 314–315.

Korona, R., 1999 Unpredictable fitness transitions between haploid and diploid strains of the genetically loaded yeast *Saccharomyces cerevisiae*. Genetics **151**: 77–85.

Kuranda, K., V. Leberre, S. Sokol, G. Palamarczyk, and J. Francois, 2006 Investigating the caffeine effects in the yeast *Saccharomyces cerevisiae* brings new insights into the connection between TOR, PKC and Ras/cAMP signalling pathways. Mol. Microbiol. **61**: 1147–1166.

Lai, C., R. F. Lyman, A. D. Long, C. H. Langley, and T. F. Mackay, 1994 Naturally occurring variation in bristle number and DNA polymorphisms at the scabrous locus of *Drosophila melanogaster.* Science **266**: 1697–1702.

Lamb, T. M., W. Xu, A. Diamond, and A. P. Mitchell, 2001 Alkaline response genes of *Saccharomyces cerevisiae* and their relationship to the RIM101 pathway. J. Biol. Chem. **276**: 1850–6.

Lee, D. H., and B. O. Palsson, 2010 Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-Propanediol. Appl Environ Microb **76**: 4158–4168.

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.

Lenski, R., M. Rose, S. Simpson, and S. Tadler, 1991 Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. Am Nat **138**: 1315–1341.

Lewis Jr, W., 1985 Nutrient scarcity as an evolutionary cause of haploidy. Am Nat **125**: 692–701.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, *et al.*, 2009 The sequence alignment/map format and samtools. Bioinformatics **25**: 2078–2079.

Lindsley, D. L., L. Sandler, B. S. Baker, A. T. Carpenter, R. E. Denell, *et al.*, 1972 Segmental aneuploidy and the genetic gross structure of the Drosophila genome. Genetics **71**: 157–184.

Losos, J. B., 1992 The evolution of convergent structure in Caribbean Anolis communities. Syst Biol **41**: 403–420.

Losos, J. B., 2011 Convergence, adaptation, and constraint. Evolution **65**: 1827–1840.

Luria, S., and M. Delbrück, 1943 Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28**: 491.

Lynch, M., W. Sung, K. Morris, N. Coffey, and C. Landry, 2008 A genome-wide view of the spectrum of spontaneous mutations in yeast. PNAS **105**.

Mable, B., 2001 Ploidy evolution in the yeast *Saccharomyces cerevisiae*: a test of the nutrient limitation hypothesis. J Evol Biol **14**: 157–170.

Mable, B., and S. Otto, 2001 Masking and purging mutations following EMS treatment in haploid, diploid and tetraploid yeast *Saccharomyces cerevisiae*. Genet Res **77**: 9–26.

MacLean, R. C., and A. Buckling, 2009 The distribution of fitness effects of beneficial mutations in *Pseudomonas aeruginosa*. PLoS Genet **5**: e1000406.

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. Philos T R Soc B **365**: 2439–2450.

Mandegar, M., and S. Otto, 2007 Mitotic recombination counteracts the benefits of genetic segregation. Proc Roy Soc B **274**: 1301–1307.

Matsumoto, T., A. Ellsmore, S. Cessna, and P. Low, 2002 An osmotically induced cytosolic Ca2+ transient activates calcineurin signaling to mediate ion homeostasis and salt tolerance of *Saccharomyces cerevisiae*. J Biol Chem **277**: 33075–33080.

Mcbride, R., D. Greig, and M. Travisano, 2008 Fungal viral mutualism moderated by ploidy. Evolution **62**: 2372–2380.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature **351**: 652–654.

Meiklejohn, C. D., and D. L. Hartl, 2002 A single mode of canalization. TREE **17**: 468–473.

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.

M'Gonigle, L. K., and S. P. Otto, 2011 Ploidy and the evolution of parasitism. Proc Roy Soc B **278**: 2814–2822.

Minet, M., A. Grossenbacher-Grunder, and P. Thuriaux, 1980 The origin of a centromere effect on mitotic recombination. a study in the fission yeast *Schizosaccharomyces pombe*. Curr Genet **2**: 53–60.

Minty, J. J., A. A. Lesnefsky, F. Lin, Y. Chen, T. A. Zaroff, *et al.*, 2011 Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*. Microb Cell Fact **10**: 18.

Mongold, J., and R. E. Lenski, 1996 Experimental rejection of a nonadaptive explanation for increased cell size in *Escherichia coli*. J Bacteriol **178**: 5333–5334.

Mukai, T., S. Chigusa, L. Mettler, and J. Crow, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. Genetics **72**: 335–355.

Nass, R., K. Cunningham, and R. Rao, 1997 Intracellular sequestration of sodium by a novel Na+/H+ exchanger in yeast is enhanced by mutations in the plasma membrane H+-ATPase. J Biol Chem **272**: 26145–26152.

Nei, M., 1987 *Molecular evolutionary genetics*. Columbia University Press.

Nes, W., P. Jayasimha, W. Zhou, R. Kanagasabai, C. Jin, *et al.*, 2004 Sterol methyltransferase: Functional analysis of highly conserved residues by site-directed mutagenesis. Biochemistry **43**: 569–576.

Nevoigt, E., and U. Stahl, 1997 Osmoregulation and glycerol metabolism in the yeast *Saccharomyces cerevisiae*. FEMS Microbiol Rev **21**: 231–241.

Nishant, K., W. Wei, E. Mancera, J. Argueso, A. Schlattl, *et al.*, 2010 The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. PLoS. Genet. **6**: 5–10.

Novak, M., T. Pfeiffer, R. Lenski, U. Sauer, and S. Bonhoeffer, 2006 Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. Am Nat **168**: 242–251.

Ohnishi, G., K. Endo, A. Doi, A. Fujita, and Y. Daigaku, 2004 Spontaneous mutagenesis in haploid and diploid *Saccharomyces cerevisiae*. Biochem Bioph Res Co **325**: 928–933.

Ohnishi, O., 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in Drosophila. II. Homozygous effect of polygenic mutation. Genetics **87**: 529–545.

Orr, H., and S. Otto, 1994 Does diploidy increase the rate of adaptation? Genetics **136**: 1475–1480.

Orr, H. A., 2003 The distribution of fitness effects among beneficial mutations. Genetics **163**: 1519–1526.

Orr, H. A., 2005 The genetic theory of adaptation: a brief history. Nat Rev Genet **6**: 119–127.

Orr, H. A., and R. L. Unckless, 2008 Population extinction and the genetics of adaptation. Am. Nat. **172**: 160–169.

Ostrowski, E., D. Rozen, and R. Lenski, 2005 Pleiotropic effects of beneficial mutations in *Escherichia coli*. Evolution **59**: 2343–2352.

Ostrowski, E., R. Woods, and R. Lenski, 2008 The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. Proc Roy Soc B **275**: 277–284.

Otto, S. P., and D. B. Goldstein, 1992 Recombination and the evolution of diploidy. Genetics **131**: 745–751.

Otto, S. P., and J. Whitton, 2000 Polyploid incidence and evolution. Annu Rev Genet **34**: 401–437.

Palacios, J., and R. Serrano, 1978 Proton permeability induced by polyene antibiotics - plausible mechanism for their inhibition of maltose fermentation in yeast. FEBS Lett **91**: 198–201.

Pandit, M., and M. Pocock, 2011 Ploidy influences rarity and invasiveness in plants. J Ecol : 1108–1115.

Paquin, C., and J. Adams, 1983 Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. Nature **302**: 495–500.

Parfrey, L. W., D. J. G. Lahr, and L. A. Katz, 2008 The dynamic nature of eukaryotic genomes. Mol Biol Evol **25**: 787–794.

Parks, L. W., and W. M. Casey, 1995 Physiological implications of sterol biosynthesis in yeast. Annu Rev Microbiol **49**: 95–116.

Parreiras, L. S., L. M. Kohn, and J. B. Anderson, 2011 Cellular Effects and Epistasis among Three Determinants of Adaptation in Experimental Populations of *Saccharomyces cerevisiae*. Euk Cell **10**: 1348–1356.

Pelosi, L., L. Kühn, D. Guetta, J. Garin, J. Geiselmann, *et al.*, 2006 Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. Genetics **173**: 1851–1869.

Peña, M., J. Lee, and D. Thiele, 1999 A delicate balance: homeostatic control of copper uptake and distribution. J Nutr **129**: 1251–1260.

Pepin, K. M., and H. A. Wichman, 2008 Experimental evolution and genome sequencing reveal variation in levels of clonal interference in large populations of bacteriophage ϕx174. BMC Evol Biol **8**: 85.

Peters, A., D. Halligan, M. Whitlock, and P. Keightley, 2003 Dominance and overdominance of mildly deleterious induced mutations for fitness traits in *Caenorhabditis elegans*. Genetics **165**: 589–599.

Poole, A. M., M. J. Phillips, and D. Penny, 2003 Prokaryote and eukaryote evolvability. Biosystems **69**: 163–185.

Pörtner, H., A. Bennett, F. Bozinovic, A. Clarke, M. Lardies, *et al.*, 2006 Trade-offs in thermal adaptation: the need for a molecular to ecological integration. Physiol Biochem Zoo **79**: 295–313.

R Development Core Team, 2008 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

R Development Core Team, 2011 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rokyta, D. R., C. J. Beisel, P. Joyce, M. T. Ferris, C. L. Burch, *et al.*, 2008 Beneficial fitness effects are not exponential for two viruses. J Mol Evol **67**: 368–376.

Rokyta, D. R., P. Joyce, S. B. Caudle, and H. A. Wichman, 2005 An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat Genet **37**: 441–444.

Rouzine, I. M., E. Brunet, and C. O. Wilke, 2008 The traveling-wave approach to asexual evolution: Muller's ratchet and speed of adaptation. Theor Pop Biol g **73**: 24–46.

Rozen, D., J. de Visser, and P. Gerrish, 2002 Fitness effects of fixed beneficial mutations in microbial populations. Curr Biol **12**: 1040–1045.

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

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

Schluter, D., E. A. Clifford, M. Nemethy, and J. S. McKinnon, 2004 Parallel evolution and inheritance of quantitative traits. Am Nat **163**: 809–822.

Schoustra, S. E., A. J. M. Debets, M. Slakhorst, and R. F. Hoekstra, 2006 Reducing the cost of resistance; experimental evolution in the filamentous fungus *Aspergillus nidulans*. J Evol Biol **19**: 1115–1127.

Segrè, A. V., A. W. Murray, and J.-Y. Leu, 2006 High-resolution mutation mapping reveals parallel experimental evolution in yeast. Plos Biol **4**: e256.

Selk, E., and C. Wills, 1998 Mismatch repair and the accumulation of deleterious mutations influence the competitive advantage of MAT (mating type) heterozygosity in the yeast *Saccharomyces cerevisiae*. Genet Res **71**: 1–10.

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 Genet **5**: e1000705.

Serrano, R., H. Martin, A. Casamayor, and J. Arino, 2006 Signaling alkaline pH stress in the yeast *Saccharomyces cerevisiae* through the Wsc1 cell surface sensor and the Slt2 MAPK pathway. J Biol Chem **281**: 39785–39795.

Sia, R., B. Urbonas, and E. Sia, 2003 Effects of ploidy, growth conditions and the mitochondrial nucleoid-associated protein Ilv5p on the rate of mutation of mitochondrial DNA in *Saccharomyces cerevisiae*. Curr Genet **44**: 26–37.

Simmons, M. J., and J. F. Crow, 1977 Mutations affecting fitness in Drosophila populations. Annu Rev Genetics **11**: 49–78.

Slater, M., 1973 Effect of reversible inhibition of deoxyribonucleic acid synthesis on the yeast cell cycle. J Bacteriol **113**: 263–270.

St Charles, J., E. Hazkani-Covo, Y. Yin, S. L. Andersen, F. S. Dietrich, *et al.*, 2012 High-resolution genome-wide analysis of irradiated (UV and Γ-rays) diploid yeast cells reveals a high frequency of genomic loss of heterozygosity (LOH) events. Genetics **190**: 1267–1284.

Stahl, G., S. Salem, L. Chen, B. Zhao, and P. Farabaugh, 2004 Translational accuracy during exponential, postdiauxic, and stationary growth phases in *Saccharomyces cerevisiae*. Eukaryot Cell **3**: 331–338.

Szafraniec, K., R. H. Borts, and R. Korona, 2001 Environmental stress and mutational load in diploid strains of the yeast *Saccharomyces cerevisiae*. PNAS **98**: 1107–1112.

Szafraniec, K., D. Wloch, P. Sliwa, and R. Borts, 2003 Small fitness effects and weak genetic interactions between deleterious mutations in heterozygous loci of the yeast *Saccharomyces cerevisiae*. Genet Res **82**: 19–31.

Temina, A., I. Tolstorukov, V. Korogodin, and A. Golobov, 1979 Comparative study of the protein makeup in diploid and haploid forms of *Saccharomyces* and *Pichia*. Mikrobiologiia **48**: 610–616.

Tenaillon, O., A. Rodriguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, *et al.*, 2012 The molecular diversity of adaptive convergence. Science **335**: 457–461.

Thompson, D., M. Desai, and A. Murray, 2006 Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. Curr Biol **16**: 1581–1590.

Toprak, E., A. Veres, J.-B. Michel, R. Chait, D. L. Hartl, *et al.*, 2011 Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. Nat Genet **44**: 101–105.

Torres, E. M., T. Sokolsky, C. M. Tucker, L. Y. Chan, M. Boselli, *et al.*, 2007 Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science **317**: 916–24.

Vassilieva, L., A. Hook, and M. Lynch, 2000 The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. Evolution **54**: 1234–1246.

Wang, R. L., A. Stec, J. Hey, L. Lukens, and J. Doebley, 1999 The limits of selection during maize domestication. Nature **398**: 236–239.

Weinreich, D., R. Watson, and L. Chao, 2005 Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution **59**: 1165–1174.

Weiss, R., J. Kukora, and J. Adams, 1975 The relationship between enzyme activity, cell geometry, and fitness in *Saccharomyces cerevisiae*. PNAS **72**: 794–798.

Winzeler, E. A., 1999 Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science **285**: 901–906.

Wong, A., and R. Kassen, 2011 Parallel evolution and local differentiation in quinolone resistance in *Pseudomonas aeruginosa*. Microbiol **157**: 937–944.

Woods, R., 1971 Nystatin-resistant mutants of yeast: alterations in sterol content. J Bacteriol **108**: 69–73.

Woods, R., D. Schneider, C. L. Winkworth, M. A. Riley, and R. E. Lenski, 2006 Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. PNAS **103**: 9107–9112.

Wright, S., 1934 Molecular and evolutionary theories of dominance. Am Nat **63**: 24–53.

Wu, C.-Y., P. A. Rolfe, D. K. Gifford, and G. R. Fink, 2010 Control of transcription by cell size. PLoS Biol **8**: e1000523.

Zeyl, C., and J. DeVisser, 2001 Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. Genetics **157**: 53–61.

Zeyl, C., T. Vanderford, and M. Carter, 2003 An evolutionary advantage of haploidy in large yeast populations. Science **299**: 555–8.

Appendix A

Appendix for Chapter 2: Cryptic fitness advantage

A.1 Supporting Tables

Table A.1: Cell volume correlates strongly with surface area, eccentricity and surface area:volume across colonies isolated across the timeseries

	r	statistic
volume & surface area	0.997	t_{35}=73.5, p<0.00001
volume & eccentricity	0.627	t_{35}=4.76, p<0.00001
volume & surface area:volume	-0.978	t_{35}=-27.5, p<0.00001

Table A.2: Cell volume correlates strongly with surface area, eccentricity and surface area:volume across colonies isolated at generations 1023 & 1302

	r	statistic
volume & surface area	0.999	t_{21}=108.6, p<0.00001
volume & eccentricity	0.857	t_{21}=7.60, p<0.00001
volume & surface area:volume	-0.991	t_{21}=-34.5, p<0.00001

A.2 Supporting Figures

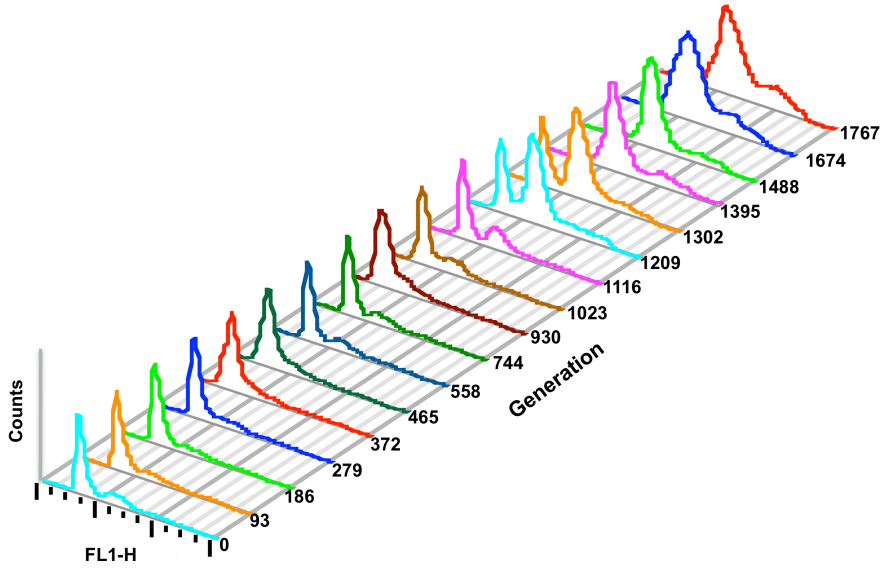


Figure A.1: Freezer culture frozen down from the initial evolution experiment was inoculated straight into 10 mL of YPD and grown for 48 hours. We then used hydroxyurea to synchronize the cell cycle and measured 30 000 cells each time point. This assay provides us with a snapshot of ploidy transition from a haploid population at generation 0 to a diploid population after generation 1395. Throughout, there is a second smaller peak at double the current ploidy level due to some cells remaining in the G2 phase (see Figure A.3).

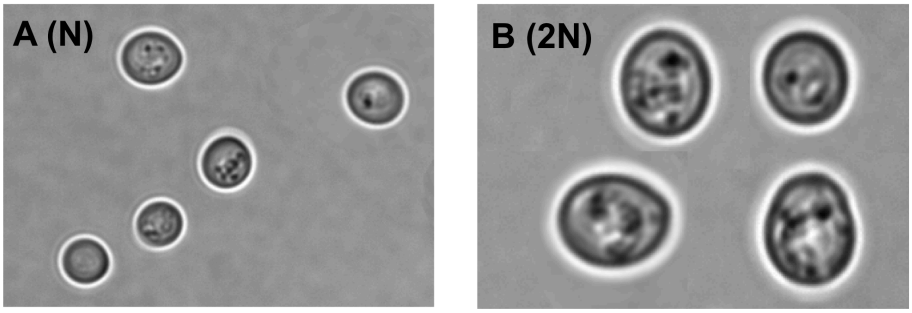


Figure A.2: Representative images of haploid (A) and diploid (B) cells used in imaging experiment. Elipses were manually drawn around cells to measure the major and minor axes for use in volume, surface area and eccentricity calculations.

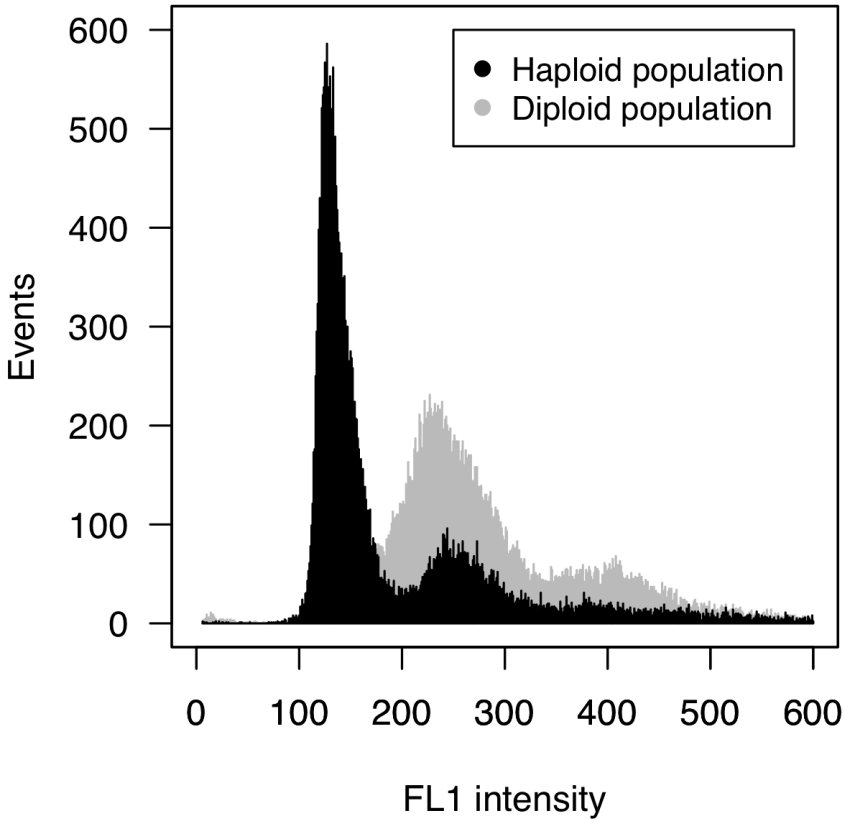


Figure A.3: Hydroxyurea is used to synchronize the cell cycle of populations. Presented are the measurement of 30 000 cells from a population composed entirely of haploids (black) and 30 000 cells from a population of diploids (grey). This method is not perfect, as some cells escape arrest. We have found the fraction of un-arrested cells to be fairly consistent, however, and as we focus our results on the difference between the ratio of haploids/diploids from one time point to another, this should not bias our conclusions.

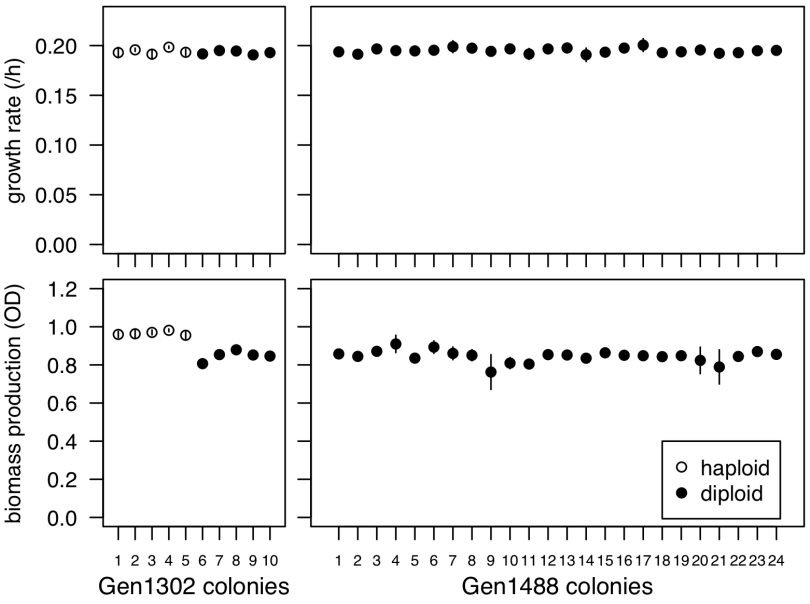


Figure A.4: Fitness components of 1488 generation diploids do not predict diploid advantage. Growth rate and biomass production from 10 colonies (5 haploid and 5 diploid) isolated at 1302 generations and 24 diploid colonies isolated from 1488 generations were measured on a Bioscreen C Microbiology Workstation (Thermo Labsystems). Although only diploid colonies were present at 1550 generations, these fitness components do not predict a diploid advantage over the haploid colonies (first panel) that were present immediately before diploid takeover.

Appendix B

Appendix for Chapter 3: Haploids adapt faster than diploids across a range of environments

B.1 Experimental Environments

The six stressful environments were created by the addition of stressor to a YPD base. NaCl, HCl (1M), and KOH (5M) were added directly to YPD before autoclaving while the appropriate amount of stock solutions of caffeine (0.103M), ethanol (95%), and nystatin (1mM) were added to YPD after autoclaving. To ensure there was no difference in the concentration of nutrients (YPD) or stressors, the volume of all bottles was kept constant through the addition of sterilized water as required, post-autoclave. All medium was prepared in batches weekly. Midway through the experiment nystatin was added individually to each YPD+nystatin 10mL test tube to minimize the risk of contamination.

B.2 Nystatin Competitions

Using the same protocol as with the other environments (50μL reference to 50μL competing strains), we initially found that the fraction of non-fluorescing cells (experimental strains) in many cases reached 95% by day 2 (the second measurement), even in the strains isolated from generation 49. To give us more power to detect selection, we thus decreased their starting volume (to 25μL experimental strain and 75μL reference strain) and added an extra measurement day (measured on days 0, 2, 3 and 4).

B.3 Comparing the Rate of Haploid to Diploid Adaptation

We use equation A2 from Otto and Whitton (2000) to interpret the rate of adaptation in asexual populations in terms of the selection coefficients underlying the adaptation. This method is based on early work by Kimura and Crow (1964), who noted that for a mutation to fix in an asexual population, it must occur within a lineage already carrying any other beneficial mutations that are destined to fix. As described by Otto and Whitton (2000), this logic can be used to determine the rate at which fitness rises over time in an asexual population with ploidy level c as the inverse of the number of generations that pass on average between the appearance of two successful beneficial mutations (σ_c and σ_c'), where success is defined as the mutation ultimately becoming fixed within the population.

$$\triangle W_{asexual} = \frac{\sigma_c \sigma_c'}{ln[cN(Exp[\frac{\sigma_c'}{2c\nu N\sigma_c}] - 1)(\frac{\sigma_c'}{\sigma_c' + \sigma_c})]} \tag{B.1}$$

(equation B.1 corrects typographical errors in the original Otto and Whitton 2000 paper).

Equation B.1 describes the long-term average rate of fitness increase; we assume here that the fitness changes over the 140 generations of our experimental treatments can be used as a proxy for $\triangle$W, in the absence of more detailed information about the genetic changes that have occurred.

To estimate selection, we assume that the beneficial alleles that are destined to fix have a roughly constant advantageous effect size over the time course of these experiments ($\sigma_c = \sigma_c' = s$ for haploids, sh for diploids). We allow haploid and diploid populations to have different effective population sizes ($N_{e,h}$ and $N_{e,d}$, respectively) and different beneficial mutation rates (ν_h and ν_d). We can then describe both haploid (equation S2, with c=1) and diploid (equation S3, with c=2) rates of adaptation:

$$\triangle W_h = \frac{s^2}{ln[N_{e,h}(Exp[\frac{s}{2\nu_h N_{e,h} s}] - 1)(\frac{1}{2})]} \tag{B.2}$$

$$\triangle W_d = \frac{s^2 h^2}{ln[2N_{e,d}(Exp[\frac{sh}{4\nu_d N_{e,d} sh}] - 1)(\frac{1}{2})]} \tag{B.3}$$

Assuming that s in haploids equals that in diploids and taking the ratio of the rate of haploid adaptation (equation S2) and diploid adaptation (equation S3), the rate of change in fitness can be used to obtain a dominance coefficient of beneficial mutations:

$$h = \sqrt{\frac{ln[2N_{e,d}(Exp[\frac{1}{4\nu_d N_{e,d}}] - 1]) \cdot \triangle W_d}{ln[N_{e,h}(Exp[\frac{1}{2\nu_h N_{e,h}}] - 1) \cdot \triangle W_h}} \tag{B.4}$$

In the text, we assumed an equal mutation rate of (10^{-7}) for haploids and diploids. The inferred dominance coefficients were not, however, sensitive to the mutation rate across a broad range of potential values (Figure B.3). The inferences were also unaffected if the measured genomic mutation rates were used to scale the relative rate of mutations in haploids and diploids to $\nu_h/\nu_d = 3.3/2.9$ (Lynch *et al.*, 2008; Nishant *et al.*, 2010). Only if the haploid mutation rate were orders of magnitude smaller than assumed in the text would our dominance estimates have been overestimated (Figure B.4).

Equations (S2) – (S4) do not account for there being a distribution of selective effects or for the fact that only the best of the beneficial mutations that arise are likely to fix within the population. That is, competition among beneficial mutations for fixation (clonal interference) will lead to the fixation of mutations with a higher selective advantage s (high hs in diploids) than expected based on the average of all possible beneficial mutations (Gerrish and Lenski, 1998; Rozen *et al.*, 2002). These equations also assume that beneficial mutations destined to fix are nested within the previous lineage destined to fix. With high enough mutation rates and population sizes, however, leap-frogging becomes possible, such that multiple beneficial mutations can arise and change the fate of a previously doomed lineage. To investigate the impact of this possibility, we also applied equation (52) from Rouzine *et al.* (2008), which calculates the speed of a travelling wave of adaptation and accounts for stochasticity at the wave front; this theory allows for multiple mutations to rescue genotypes of lower fitness. Similar selection and dominance coefficients were estimated by this method (Table B.3).

B.4 Confidence Intervals on h

To obtain 95% confidence intervals for the dominance coefficient, we bootstrapped 10000 sets of five haploid and five diploid rates of adaptation from a normal distribution (with means and standard deviations equal to the means and standard deviations of our measured results for haploids and diploids in each environment). Effective population sizes here are very large, and mutations are not limiting; there is virtually no difference in dominance estimate whether ancestral or evolved population sizes are used, even in YPD+NaCl where the population size significantly decreased over the experiment (results not shown). We thus bootstraped 10000 sets of five haploid and five diploid effective population sizes with mean and standard deviation equal to the mean and standard deviation of the effective population sizes in the original experiment (after averaging ancestral and evolved measurements). The bootstrapped datasets were then used to calculate dominance 10000 times for a particular environment. The upper and lower bounds were set to the 97.5 and 2.5 quantiles from the bootstrap distribution of dominance coefficients and represent confidence intervals.

In a number of cases the bootstrap procedure led to negative estimates of the rate of adaptation (primarily in estimating diploid rates of adaptation in YPD, YPD+HCl and YPD+ethanol, though also a small number of times for both ploidy levels in other environments); in these situation the inferred h value from equation (S4) would be complex. Because the population sizes were large, we assumed that negative rates of adaptation were due to sampling error, and we forced the rate of adaptation to be very small but positive (10^{-6}, though results were insensitive to forced rates between 10^{-4} – 10^{-9}).

B.5 Only Single Mutations are Likely Present at High Frequency in Most Lines

Here, we ask what the minimum time to reach 50% would be for mutations of varying beneficial effects (s) and dominance coefficients (h). To do so, we use theoretical results from Campos and Wahl (2009) developed for these types of evolutionary experiments with periodic bottlenecks, calculating $T_{50\%} \approx \frac{1}{2}\frac{ln(N_0)}{hs_b}$ (Campos &Wahl 2009, equation 6 and Supplementary material). For a dominant mutation (h=1) to reach 50% in 200 generations, the selective advantage (s) must be at least 0.13 (Figure B.2). As the dominance of the beneficial mutation decreases, the effect size of the mutations must correspondingly increase to reach 50% within 200 generations (Figure B.2). The results presented use the average population size transfered daily (N_0) across all environments, though population size did not greatly affect the rate at which the beneficial mutations are predicted to reach 50% frequency. Populations an order of magnitude larger of smaller than our measured population sizes (Figure B.2) show nearly identical results. We thus believe that it is likely only single mutations are present at high frequency in any of our lines.

Table B.1: Two Way ANOVA results for effective population sizes. Significant factors are shown in bold.

	Ploidy	Time	Ploidy * Time
YPD	$F_{1,16}$ = 46.1 **p< 0.0001**	$F_{1,16}$ = 1.9 p > 0.05	$F_{1,16}$ = 0.053 p > 0.05
YPD + HCl	$F_{1,16}$ = 11.1 **p= 0.004**	$F_{1,16}$ = 0.2 p > 0.05	$F_{1,16}$ = 2.5 p > 0.05
YPD + Ethanol	$F_{1,16}$ = 62.8 **p< 0.001**	$F_{1,16}$ = 0.1 p > 0.05	$F_{1,16}$ = 3.7 p > 0.05
YPD + KOH	$F_{1,16}$ = 143.6 **p< 0.0001**	$F_{1,16}$ = 0.5 p > 0.05	$F_{1,16}$ = 1.0 p > 0.05
YPD + Nystatin	$F_{1,16}$ = 13.2 **p= 0.002**	$F_{1,16}$ = 0.1 p > 0.05	$F_{1,16}$ = 3.3 p > 0.05
YPD + NaCl	$F_{1,15}$ = 25.7 **p< 0.0001**	$F_{1,15}$ = 38.2 **p< 0.0001**	$F_{1,15}$ = 12.3 **p= 0.003**
YPD + Caffeine	$F_{1,16}$ = 5.5 **p = 0.03**	$F_{1,16}$ = 0.2 p > 0.05	$F_{1,16}$ = 0.6 p > 0.05

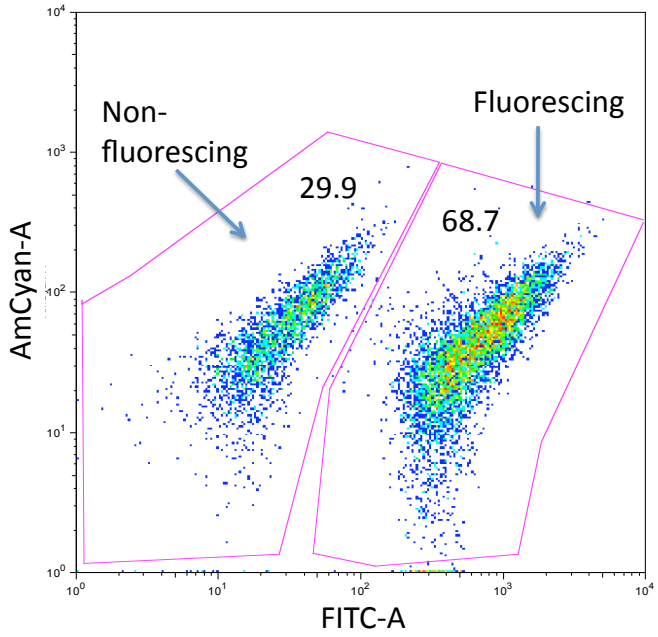


Figure B.1: 10000 cells from each culture of interest were read in 96 well plates on an LSRII. Each cell is plotted on AmCyan-A and FITC-A axes which separates out non-fluorescing (left gate) from fluorescing cells (right gate). Numbers in each gate indicate the proportion of cells; this number is used to determine the fraction of non-fluorescing cells for further analysis (e.g., $NonFluor = \frac{29.9}{29.9+68.7}$ in this case).

Table B.2: Rate of adaptation between generations 47 and 187 within each lineage. These data were divided by 140 generations to obtain the per generation rate used in Figure 1.

	Haploid lines ($140\times\Delta W_h$)	Diploid lines ($140\times\Delta W_d$)
YPD	0.098	-0.015
	0.022	-0.011
	0.021	-0.003
	0.018	0.008
	NA*	-0.012
YPD + HCl	-0.003	0.021
	0.026	-0.06
	0.047	0.014
	0.063	0.010
	0.042	0.014
YPD + Ethanol	0.028	0.001
	0.029	-0.041
	0.045	-0.019
	0.038	-0.005
	0.019	0.009
YPD + KOH	0.043	0.023
	0.028	0.023
	0.024	0.048
	0.043	0.025
	0.075	0.033
YPD + Nystatin	0.048	0.030
	0.070	0.077
	0.064	0.051
	0.079	0.053
	0.0089	0.040
YPD + NaCl	0.117	0.036
	0.117	0.069
	0.123	0.051
	0.088	0.035
	0.100	0.054
YPD + Caffeine	0.176	0.109
	0.177	0.071
	0.255	0.069
	0.086	0.060
	0.077	0.186

* One haploid line in YPD became contaminated over the course of the experiment, and data was not collected.

Table B.3: Selection and dominance coefficients in haploids and diploids based on travelling wave theory of Rouzine et al. (2008). Estimates of s and hs are based on equation (52), with ν set to 10^{-7} using the averaged N_e values. Estimates of s and hs are somewhat sensitive to the mutation rate assumed, but the dominance coefficient h is robust and similar to that shown in Figure 3.

	Haploid lines (s)	Diploid lines (hs)	Dominance (h)
YPD s	0.022	0*	0*
YPD + HCl	0.048	0*	0*
YPD + Ethanol	0.046	0*	0*
YPD + KOH	0.055	0.046	0.83
YPD + Nystatin	0.074	0.057	0.76
YPD + NaCl	0.091	0.059	0.65
YPD + Caffeine	0.108	0.085	0.79

* The average measured rate of adaptation of diploids lines was negative in these environments.

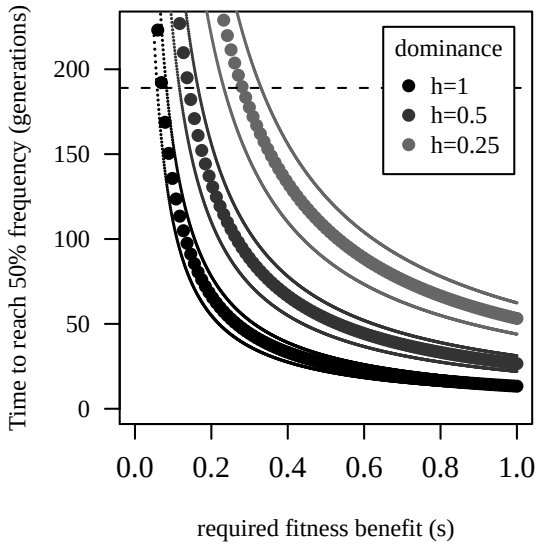


Figure B.2: Time required for a beneficial mutation to reach 50% frequency in our experiments with a starting population size of 595067 (the measured average population size transfered daily across all environments; N_0 as in Campos and Wahl 2009 equation 6) and with periodic bottlenecks every 6.7 generations. Lines around the main points indicate the result found when effective population sizes are an order of magnitude larger (upper bound) or smaller (lower bound).

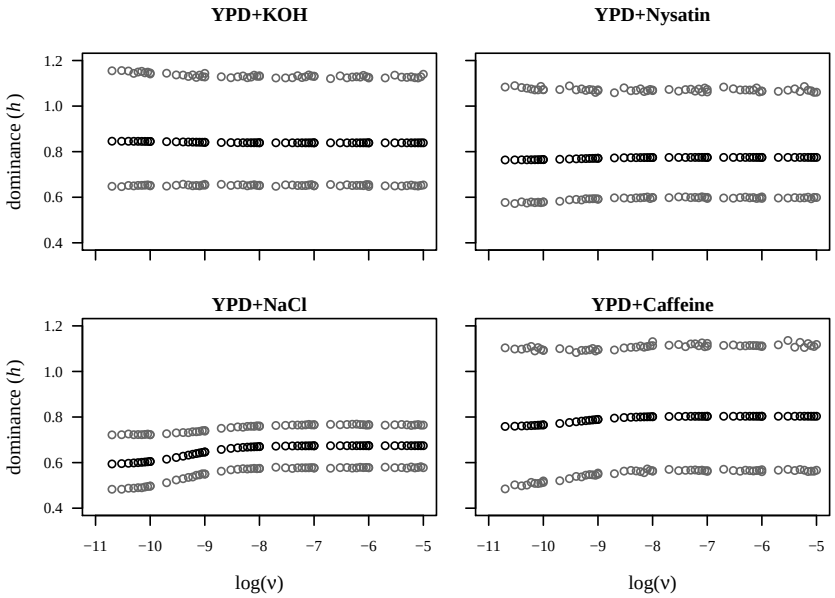


Figure B.3: Dominance estimates are not sensitive to changing ν, keeping the haploid mutation rate equal to the diploid mutation rate. Black dots are based on equation (S4), using the mean rate of adaptation and mean effective population sizes observed in haploids and in diploids. Grey dots indicate 95% confidence intervals obtained by simultaneous parametric bootstrapping haploid and diploid rates of adaptation and haploid and diploid effective population sizes (as in Figure 3) while changing the mutation rate (ν).

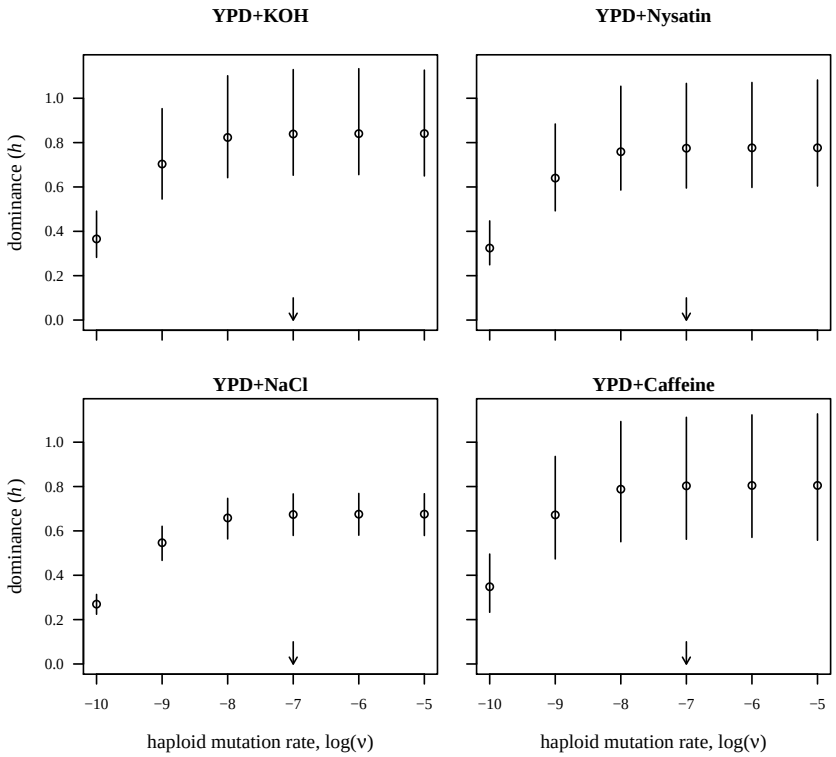


Figure B.4: The sensitivity of dominance estimates to decreasing the haploid mutation rate. The diploid mutation rate was 10^{-7} (arrow) as in the simulations presented above, while the haploid mutation rate was changed. If the haploid mutation rate was three orders of magnitude lower than for diploids, the dominance estimates presented in Figure 3 would be overestimated.

Appendix C

Appendix for Chapter 4 : Parallel genetic change and non-parallel gene-environment interactions underlie the first step of nystatin

C.1 Appearance of Identical Mutations

Five specific ergosterol mutations were sampled from multiple lines (Table 1). The most likely explanation is that these mutations arose during population expansion before the lineages were isolated from one another and before the stressor was applied. The culture used to seed mutation acquisition screens was derived from a single wild type colony grown up overnight in YPD, an unstressful environment. Given the size of this overnight culture ($\sim 1.2 \times 10^9$ cells), there would have been approximately 30 generations of growth during this YPD phase ($2^{30} = 1.1 \times 10^9$). Despite the bottleneck to a single colony-forming unit, the population size of the source population, a plausible per-base pair mutation rate (0.33×10^{-9}, Lynch *et al.* (2008)), the hundreds of one-step mutations that could potentially result in nystatin tolerance ($\sim$ 350 different mutations based on our data of what types of mutations confer tolerance to nystatin, see section below), and the number of founding lineages (60 and 180 in screens 'a' and 'b', respectively), we calculated that there is a high probability that some beneficial mutations were segregating in the precursor population (Appendix C.2). Standing genetic variation from a colony grown for a single overnight in YPD has previously been found to play a large role in the eventual mutations that were selected in a yeast experimental evolution project (Gresham *et al.*, 2008), and we believe this is also the most likely explanation here.

Well-to-well contamination is also possible, yet unlikely. We kept track of where mutations were isolated within the 96 well plates; in no case was the same mutation isolated in neighbouring wells and in multiple cases the same mutation was isolated from different plates within the same screen. In at least the case of BMN11-15, well-to-well contamination is even less likely, as growth was seen in all wells before the first culture was isolated, thus there was little to no opportunity for contamination from one well to another (Table S2).

Although certainly possible biologically, we do not have any support for the same allelic variant arising independently in multiple lines. Mutation lines were acquired in two separate screens (denoted 'a' and 'b' in Table S2). The exact same protocol was used for both acquisition screens (see Methods), however, a different colony was grown up overnight to provide the culture used to seed all replicate wells in each screen. In no case was the same allele sampled in the two different screens.

C.2 Chance of Multiple Mutations

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

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 ~10^9cells ("source" population), from which a sample of cells is taken to establish each individual lineage (the "founding" population).

Parameters:

μ = total mutation rate to nystatin resistance (0.33 ◊ 10^{-9}(Lynch et al., 2008, PNAS); range explored: $10^{-9} - 10^{-6}$)

N1 = population size at saturation in 10mL YPD (1.2×10^9 cells, measured)

f = fraction of population sampled to found a lineage (0.001 = 10 ul/10 mL)

L = number of founding lineages started from the YPD culture (note there are two mutation accumulation "screens", the first with 60 well replicates, the second with 180)

lowerMUT = lower estimate of the target size

upperMUT = upper estimate of the target size

```
tryμ = 0.33 × 10^-9;
tryN1 = 1.2 * 10^9;
tryf = 0.001;
tryL1 = 60;
tryL2 = 180;
lowerMUT = 246;
upperMUT = 522;
```

Calculation #1: Estimating the genome-wide target size for mutations to nystatin resistance in our screens

Here, we estimate the number of sites that could potentially yield nystatin resistance by a single basepair mutation. Because multiple resistant erg3 and erg6 mutations were obtained and many of these created a stop codon, we first ask how many nucleotides in these two genes could mutate to a stop codon. We then use the observed proportion of unique mutations that were stop codons in erg3 and erg6 to correct this etimate and obtain an overall estimate of the number of sites that could potentially yield nystatin resistance. While the estimation is rough, it gives a reasonable expectation that there are hundreds of potential sites in the genome that could give rise to nystatin resistance of the nature observed in our experiment (many additional minor effect mutations are possible but were not obtained by our assay. Note that then consider how this number might be an over- or under-estimate.

Copy and paste sequences of ERG3 and ERG6 (from yeastgenome.org)

```
erg3 =
  "ATGGATTTGGTCTTAGAAGTCGCTGACCATTATGTCTTAGACGACTTGTACGCTAAAGTTCTGCCCGCTTCGTTGGCAGCTAATA\
     TTCCTGTCAAGTGGCAGAAATTGCTAGGGTTGAACAGTGGGTTCAGCAATTCTACGATTTTGCAGGAGACTTTGAACTCC\
     AAGAATGCCGTCAAAGAATGTAGAAGGTTCTACGGGCAGGTGCCATTCCTGTTTGATATGTCGACGACGTCTTTTGCATC\
     GCTATTGCCTCGTTCCAGCATCTTGAGAGAATTCCTCTCACTATGGGTTATTGTTACGATCTTTGGTTTACTACTTTACT\
     TATTCACGGCTAGTCTCAGCTACGTGTTTGTGTTTGACAAGTCGATTTTCAACCATCCTCGTTACTTGAAAAACCAAATG\
     GCAATGGAAATCAAGTTGGCAGTCAGTGCTATCCCATGGATGTCGATGTTGACCGTTCCATGGTTTGTTATGGAATTGAA\
     CGGCCATTCTAAACTATACATGAAGATTGATTATGAAAACCACGGTGTAAGGAAGCTCATTATCGAGTACTTCACTTTCA\
     TCTTTTTCACTGATTGCGGTGTGTATTTAGCGCACAGATGGTTGCATTGGCCAAGGGTCTACCGTGCTCTGCACAAGCCT\
     CATCACAAGTGGCTGGTCTGCACACCTTTCGCATCTCATTCTTTCCATCCTGTAGACGGGTTTTTGCAATCCATCTCGTA\
     CCACATCTACCCATTGATTCTGCCATTACACAAGGTTTCTTATTTGATTCTGTTCACTTTTGTTAACTTTTGGACTGTTA\
     TGATTCATGACGGTCAATACCTATCAAACAATCCTGCCGTCAACGGTACTGCCTGCCACACGGTTCACCATCTATATTTC\
     AACTACAACTACGGTCAATTCACCACTCTGTGGGACAGACTAGGGGGTTCTTACCGTAGACCAGATGACTCATTGTTTGA\
     TCCTAAGTTAAGAGATGCTAAGGAGACCTGGGACGCTCAAGTTAAGGAAGTTGAACATTTCATCAAGGAGGTCGAAGGTG\
     ATGATAATGATAGAATCTATGAAAACGACCCAAATACCAAGAAGAACAACTGA";

erg6 =
  "ATGAGTGAAACAGAATTGAGAAAAAGACAGGCCCAATTCACTAGGGAGTTACATGGTGATGATATTGGTAAAAAGACAGGTTTGA\
     GTGCATTGATGTCGAAGAACAACTCTGCCCAAAAGGAAGCCGTTCAGAAGTACTTGAGAAATTGGGATGGTAGAACCGAT\
     AAAGATGCCGAAGAACGTCGTCTTGAGGATTATAATGAAGCCACACATTCCTACTATAACGTCGTTACAGATTTCTATGA\
     ATATGGTTGGGGTTCCTCTTTCCATTTCAGCAGATTTTATAAAGGTGAGAGTTTCGCTGCCTCGATAGCAAGACATGAAC\
     ATTATTTAGCTTACAAGGCTGGTATTCAAAGAGGCGATTTAGTTCTCGACGTTGGTTGTGGTGTTGGGGGCCCAGCAAGA\
     GAGATTGCAAGATTTACCGGTTGTAACGTCATCGGTCTAAACAATAACGATTACCAAATTGCCAAGGCAAAATATTACGC\
     TAAAAAATACAATTTGAGTGACCAAATGGACTTTGTAAAGGGTGATTTCATGAAAATGGATTTCGAAGAAAACACTTTCG\
     ACAAAGTTTATGCAATTGAGGCCACATGTCACGCTCCAAAATTAGAAGGTGTATACAGCGAAATCTACAAGGTTTTGAAA\
     CCGGGTGGTACCTTTGCTGTTTACGAATGGGTAATGACTGATAAATATGACGAAAACAATCCTGAACATAGAAAGATCGC\
     TTATGAAATTGAACTAGGTGATGGTATCCCAAAGATGTTCCATGTCGACGTGGCTAGGAAAGCATTGAAGAACTGTGGTT\
     TCGAAGTCCTCGTTAGCGAAGACCTGGCGGACAATGATGATGAAATCCCTTGGTATTACCCATTAACTGGTGAGTGGAAG\
     TACGTTCAAAACTTAGCTAATTTGGCCACATTTTTCAGAACTTCTTACTTGGGTAGACAATTTACTACAGCAATGGTTAC\
     TGTAATGGAAAAATTAGGTCTAGCCCCAGAAGGTTCCAAGGAAGTTACTGCTGCTCTAGAAAATGCTGCGGTTGGTTTAG\
     TTGCCGGTGGTAAGTCCAAGTTATTCACTCCAATGATGCTTTTCGTCGCTAGGAAGCCAGAAAACGCCGAAACCCCCTCC\
     CAAACTTCCCAAGAAGCAACTCAATAA";
```

Calculate the number of codons
there are 366 codons in erg3
there are 384 codons in erg6

```
lenERG3 = Length[Characters[erg3]] / 3
```

```
366
```

```
lenERG6 = Length[Characters[erg6]] / 3
```

```
384
```

Split the strings into codons

```
codonsERG3 =
 Table[StringJoin[Table[Characters[erg3][[j]], {j, 3 * i - 2, 3 * i}]], {i, 1, lenERG3}]
codonsERG6 =
 Table[StringJoin[Table[Characters[erg6][[j]], {j, 3 * i - 2, 3 * i}]], {i, 1, lenERG6}]
```

The stop codons are: UAA, UAG, UGA (TAA, TAG, TGA)

Thus, the following 27 codons can mutate to a stop codon by a single basepair change:

```
{"CAA", "GAA", "AAA", "TGA", "TCA", "TTA", "TAG", "TAC", "TAT",
  "CAG", "GAG", "AAG", "TGG", "TCG", "TTG", "TAA", "TAC", "TAT",
  "CGA", "GGA", "AGA", "TAA", "TCA", "TTA", "TGG", "TGC", "TGT"};
```

We can exclude 4 cases involving stop codons:

```
{"CAA", "GAA", "AAA", "TCA", "TTA", "TAC", "TAT",
  "CAG", "GAG", "AAG", "TGG", "TCG", "TTG", "TAC", "TAT",
  "CGA", "GGA", "AGA", "TCA", "TTA", "TGG", "TGC", "TGT"};
```

Among these, the following codons can be hit in two places to yield a stop codon:

```
twice = {"TTA", "TCA", "TAC", "TAT", "TGG"};
```

The following codons can be hit in one (and only one) place to yield a stop codon:

```
once = {"CAA", "GAA", "AAA", "CAG", "GAG",
    "AAG", "TCG", "TTG", "CGA", "GGA", "AGA", "TGC", "TGT"};
```

Thus, the total number of sites that can mutate in erg3 or erg6 to a stop codon is:
#157 basepairs could mutate to create a stop codon in one step from erg3
#176 basepairs could mutate to create a stop codon in one step from erg3

```
Total[StringCount[codonsERG3, once]] +
 2 * Total[StringCount[codonsERG3, twice]]
```

```
157
```

```
Total[StringCount[codonsERG6, once]] +
 2 * Total[StringCount[codonsERG6, twice]]
```

```
176
```

This gives a total number of sites in erg3 and erg6 where a single mutation could generate a stop codon:

```
total = % + %%
```

```
333
```

Some stop codons, however, might not yield a nystatin-resistant phenotype, i.e., those near the end of the gene.
In our dataset, the last sampled stop codon in erg6 was at amino acid 223 and the last sampled erg3 stop was at amino acid 299.
Limiting our counts of potential stop codon hits to this point yields:

```
Total[StringCount[Take[codonsERG3, 299], once]] +
 2 * Total[StringCount[Take[codonsERG3, 299], twice]]
```

```
126
```

```
Total[StringCount[Take[codonsERG6, 223], once]] +
 2 * Total[StringCount[Take[codonsERG6, 223], twice]]
```

```
109
```

This gives a total number of sites in erg3 and erg6 where a single mutation could generate a stop codon at a position before or at the last non-sense mutation observed:

```
total = % + %%
```

```
235
```

Thus, there are at likely to be at least 235 possible one-step stop codons in erg3 & erg6 that could confer the tolerance phenotype.

In our data set, 9 out of 20 unique mutations we found were stop codons in erg3 or erg69. If we use this fraction also for the unobserved mutations, we obtain an estimate of:

```
235 / (9 / 20) // N
```

```
522.222
```

The above suggests that the genome-wide target size for mutations that would generate sufficient nystatin resistance to allow growth in our assays is somewhere in the range of 246 - 522 (the lower bound set by assuming that the only hits that we missed were other one-step stop codons, plus previous hits: 235+11; the upper bound set by assuming that we would have the same fraction of one-step stop codons in the unobserved mutations as we did in the observed mutations).

Calculation #2: Chance that the source population contains no nystatin mutations

Number of cell cycles required to produce source population:

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

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

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 dividing cells (-> 8 cells would involve 7 dividing cells: one 1->2, two 2->4, and four 4->8):

```
divisions /. N1 → {4, 8}
```

$$\{3, 7\}$$

```
divisions /. N1 → tryN1 // N
```

$$1.2 \times 10^9$$

The chance that NONE of these cell divisions involved a mutant is:

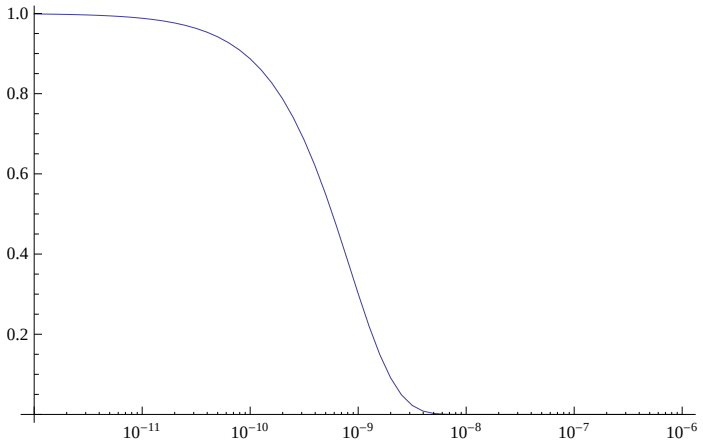
```
(1 - μ)^divisions;
```

Given that μ is small, this is approximately:

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

Assuming a population size of 1.2×10^9, this probability declines rapidly as the per-basepair mutation rate rises above 10^{-10}:

```
ListLogLinearPlot[
 Table[{10^i, nomutantsinsource[10^i, tryN1]}, {i, -12, -6, 0.1}], Joined → True]
```



Using a per-basepair mutation rate of 0.33 ◊ 10^{-9}(Lynch et al., 2008, PNAS), the chance of no mutations at a single site within the source pool would be:

```
nomutantsinsource[tryμ, tryN1]
```

$$0.673007$$

Given potentially hundreds of sites within ERG genes, the chance that none of these sites was polymoprhic in the precursor population is:

```
nomutantsinsource[tryμ × lowerMUT, tryN1]
```

4.92911×10^{-43}

```
nomutantsinsource[tryμ × upperMUT, tryN1]
```

1.68314×10^{-90}

In other words, we're virtually certain that there would be a segregating mutation at some ERG gene within the precursor population.

This does not tell us, however, how likely we are to found multiple lineages carrying the same mutation. For that, we need to determine the frequency of these mutations (many of which will have appeared in the last cell division, thus present in a single cell, and cannot found multiple lineages).

■ Calculation #3 - Calculating the fraction of the source population carrying a mutation

For a mutation rate of $0.33 \lozenge 10^{-9}$, the expected number of mutations at a specific site across all of the cells and cell divisions is approximately:

```
tryμ * divisions /. N1 → tryN1
```

0.396

The number of cell cycles required to go from one cell (original cell that lead to a colony on a plate) to 1.2×10^{9} cells (the number after letting one colony grow overnight in 10mL YPD)

```
Floor[cycles /. N1 → tryN1]
```

30

Thus, we expect less than one hit in total across all of the ~30 cell cycles at any one particular site.

Next, we derive the probability that a single mutation hits in any particular cell cycle.

At one specific site, the probability that a mutation occurs in the kth cell cycle (going from 2^{k-1} cells to 2^{k} cells) is:

```
prob[k_, μ_, N1_] = 1 - (1 - μ)^n /. n → 2^(k-1);
```

based on one minus the probability that no mutation hits. (Technically, this allows for the possibility that more than one hit would occur, but as we saw above, there is unlikely to be more than one mutation across all cell divisions and so we don't expect more than one mutation in any one cell cycle. In the following, we assume that if there is a hit in a particular cell cycle, it creates only a single daughter mutant cell.)

If a mutation does occur in the kth cycle (i.e., among the 2^{k} cells that result in this cycle, one is a new mutant), the fraction of the final population that will be mutant is:

```
frac[k_, μ_, N1_] = 1 / 2^k;
```

This gives us the probability distribution for the fraction of mutant cells in the source population (amounting to a number of mutant cells: $n = N1 / 2^{k}$), where the probability of no mutant cells at a particular site equalling:

```
1 - Sum[prob[k, μ, N1], {k, 1, cycles}]
```

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

Given L lineages started from the source population (here we use L = 80 for trial 1 and 180 for trial 2) where each lineage is started with a fraction, f, of the source population, we next calculate the probability that two or more will sample the same mutant, if that mutant consists of n cells in the source population:

```
Rehit[n_, N1_, L_, f_] = 1 - ((1 - n/N1)^(f N1))^L -
  Sum[Product[(1 - n/(N1 - (j - 1) * f * N1))^(f N1), {j, 1, i - 1}] * (1 - (1 - n/(N1 - (i - 1) * f * N1))^(f N1)) *
    Product[(1 - (n - 1)/(N1 - (j - 1) * f * N1))^(f N1), {j, i + 1, L}], {i, 1, L}]
```

$$1-\left(\left(1-\frac{n}{N1}\right)^{f\,N1}\right)^{L}-\sum_{i=1}^{L}\left(1-\left(1-\frac{n}{N1-f\,(-1+i)\,N1}\right)^{f\,N1}\right)\left(\frac{\text{Pochhammer}\left[\frac{n-N1}{f\,N1},\,-1+i\right]}{\text{Pochhammer}\left[-\frac{1}{f},\,-1+i\right]}\right)^{f\,N1}\left(\frac{\text{Pochhammer}\left[\frac{-1+n+(-1+f\,i)\,N1}{f\,N1},\,-i+L\right]}{\text{Pochhammer}\left[-\frac{1}{f}+i,\,-i+L\right]}\right)^{f\,N1}$$

This is calculated as one minus the probability that the none of the L lines get the mutant minus the probability that exactly one line gets a mutant. The latter accounts for the fact that each time a lineage is sampled, there are fewer cells in the source population and that any of the L lineages could be the one that gets the hit (sampling without replacement).

Summing over the probability distribution of when the mutation could arise we have:

```
ChanceRehit[μ_, N1_, L_, f_, m_] = Sum[prob[k, μ, N1] * Rehit[N1 / 2^k, N1, L, f], {k, 1, m}]
```

$$\sum_{k=1}^{m}\left(1-(1-\mu)^{2^{-1+k}}\right)\left(1-\left(\left(1-2^{-k}\right)^{f\,N1}\right)^{L}-\sum_{i=1}^{L}\left(1-\left(1-\frac{2^{-k}\,N1}{N1-f\,(-1+i)\,N1}\right)^{f\,N1}\right)\left(\frac{\text{Pochhammer}\left[\frac{-N1+2^{-k}\,N1}{f\,N1},\,-1+i\right]}{\text{Pochhammer}\left[-\frac{1}{f},\,-1+i\right]}\right)^{f\,N1}\left(\frac{\text{Pochhammer}\left[\frac{-1+2^{-k}\,N1+(-1+f\,i)\,N1}{f\,N1},\,-i+L\right]}{\text{Pochhammer}\left[-\frac{1}{f}+i,\,-i+L\right]}\right)^{f\,N1}\right)$$

- **For Trial 1 (L=60)**

For our parameters, at one specific site, the chance of seeing multiple lineages carrying the mutation in the first screen is:

```
ChanceRehit[tryμ, 2^30, tryL1, tryf, 30]
```

```
0.0128062
```

Given potentially hundreds of such sites, however, the chance that at least one of them them would lead to a multiple hit is very high (>95%):

```
1 - (1 - ChanceRehit[tryμ, 2^30, tryL1, tryf, 30])^lowerMUT
```

```
0.958025
```

```
1 - (1 - ChanceRehit[tryμ, 2^30, tryL1, tryf, 30])^upperMUT
```

```
0.998803
```

Even if we say that the 20 sites that we observed as nystatin resistance represents the entire target size, there is still a decent chance that we would see multiple hits:

```
1 - (1 - ChanceRehit[tryμ, 2^30, tryL1, tryf, 30])^20
```

```
0.227234
```

- **For Trial 2 (L=180 wells)**

For our parameters, at one specific site, the chance of seeing multiple lineages carrying the mutation in the second screen is:

```
ChanceRehit[tryμ, 2^30, tryL2, tryf, 30]
```

```
0.0316935
```

Given potentially hundreds of such sites, however, the chance that at least one of them them would lead to a multiple hit is very high (>99.9%):

```
1 - (1 - ChanceRehit[tryμ, 2^30, tryL2, tryf, 30])^lowerMUT
```

```
0.999638
```

```
1 - (1 - ChanceRehit[tryμ, 2^30, tryL2, tryf, 30])^upperMUT
```

```
1.
```

Even if we say that the 20 sites that we observed as nystatin resistance represents the entire target size, there is still a decent chance that we would see multiple hits:

```
1 - (1 - ChanceRehit[tryμ, 2^30, tryL2, tryf, 30])^20
```

```
0.474882
```

Calculation #4 - Calculating the chance that all sampled lineages would contain a mutation

We can also constrain our parameters by the observation that not all samples carry the same mutation. Here, we use the above calculations to determine the probability that all L lineages would carry the same mutational hit.

With L lineages (each started with a fraction of the source population, f), the probability that all sample a given mutation, if that mutant consists of n cells in the source population, is:

```
ALLhit[n_, N1_, L_, f_] = Product[(1 - (1 - n/N1)^(f N1)), {j, 1, L}]
```

$$\left(1-\left(1-\frac{n}{N1}\right)^{f\,N1}\right)^{L}$$

(This isn't exact, as we should adjust the fraction of mutant cells remaining due to sampling, but the answer will be close.)

Summing over the probability distribution of when the mutation could arise we have:

```
ChanceALLhit[μ_, N1_, L_, f_, m_] = Sum[prob[k, μ, N1] * ALLhit[N1 / 2^k, N1, L, f], {k, 1, m}]
```

$$\sum_{k=1}^{m}\left(1-\left(1-2^{-k}\right)^{f\,N1}\right)^{L}\left(1-(1-\mu)^{2^{-1+k}}\right)$$

- **For Trial 1 (L=60 wells)**

For our parameters, at one specific site, the chance of seeing all L=60 lineages carry a specific mutation in the first screen is:

```
ChanceALLhit[tryμ, 2^30, tryL1, tryf, 30]
```

```
0.0000587235
```

Even with hundreds of such sites, the chances that all lineages would be hit by the same mutation is slim (<5%):

```
1 - (1 - ChanceALLhit[tryμ, 2^30, tryL1, tryf, 30])^lowerMUT
```

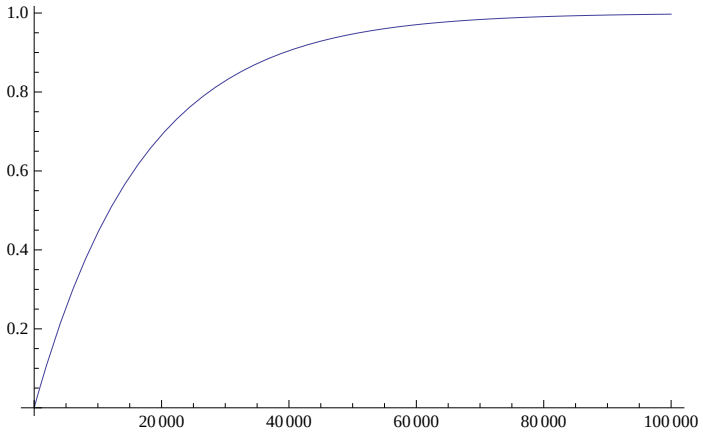
```
0.0143426
```

```
1 - (1 - ChanceALLhit[tryμ, 2^30, tryL1, tryf, 30])^upperMUT
```

```
0.0301895
```

We would, however, almost certainly sample the same hit in all L=60 lineages if the target size were >~50,000 sites:

```
Plot[1 - (1 - ChanceALLhit[tryμ, 2^30, tryL1, tryf, 30])^n, {n, 1, 100 000}]
```




```
1 - (1 - ChanceALLhit[tryμ, 2^30, tryL1, tryf, 30])^50 000
```

```
0.946936
```

Given that we did not see all of the lineages with the same hit, the above tells us that the target size could not be this large.

- **For Trial 2 (L=180 wells)**

Similarly, the chance of seeing all L=180 lineages carry a specific mutation is:

```
ChanceALLhit[tryμ, 2^30, tryL2, tryf, 30]
```

```
0.0000443119
```

Even with hundreds of such sites, the chances that all lineages would be hit by the same mutation is slim (<5%):

```
1 - (1 - ChanceALLhit[tryμ, 2^30, tryL2, tryf, 30])^lowerMUT
```

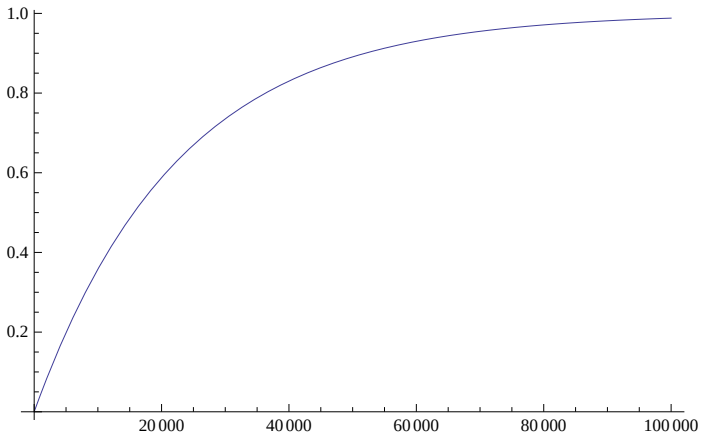
```
0.0108418
```

```
1 - (1 - ChanceALLhit[tryμ, 2^30, tryL2, tryf, 30])^upperMUT
```

```
0.0228658
```

We would, however, almost certainly sample the same hit in all L=180 lineages if the target size were >~50,000 sites:

```
Plot[1 - (1 - ChanceALLhit[tryμ, 2^30, tryL2, tryf, 30])^n, {n, 1, 100 000}]
```



```
1 - (1 - ChanceALLhit[tryμ, 2^30, tryL2, tryf, 30])^50 000
```

```
0.890917
```

Given that we did not see all of the lineages with the same hit, the above tells us that the target size could not be this large.

■ Conclusion

The above calculations inform us that there are likely to be hundreds of potential target sites leading to nystatin resistance (not tens of thousands) and that the chance that the same mutation, occurring at any one of these sites during the clonal expansion of the precursor population, would be sampled in more than one lineage is high (>95% in each screen).

C.3 Effect of Non-Ergosterol Mutations

While we attempted to minimize the number of mutations carried by the lines, secondary mutations could have arisen and fixed during the ∼30 generations of growth in YPD within the precursor population or during the ∼30 generations of growth in nystatin required for yeast precipitate to be detected. Nineteen lines carry non-synonymous nuclear mutations in genes not involved in the ergosterol pathway (Table S2). Three of these lines share the same ergosterol mutation with other BMN lines, which allows us to directly assess the phenotypic affect of these additional mutations. BMN3 has a nonsynonymous mutation in YPL039W, an uncharacterized gene. Our tolerance and fitness results from BMN3 are very similar to BMN2 and BMN4, which carry the same *ERG6* mutation but are wildtype at YPL039W, suggesting that this secondary mutation has little effect. Similarly, BMN27 carries three additional mutations: nonsynonymous mutations in YJR107W (an uncharacterized protein), *AUR1* (a protein required for sphingolipid synthesis), and an extra copy of chromosome 2. BMN27 has a higher IC_{50} in both salt and copper than the three other lines that carry the same ergosterol mutation (BMN24-26), but these differences are not significant. BMN15, with a nonsynonymous mutation in *MBP1*, also does not differ in our fitness assays from BMN11-14, lines with which it shares *ERG6* and *GDA1* mutations. The remaining lines with secondary mutations have very similar nystatin tolerance to other lines that carry mutations in the same ergosterol gene. We thus have little reason to suspect that mutations in non-ergosterol genes are strongly influencing our results.

C.4 Statistical Results Remain the Same if we Combine Lines With the Same Ergosterol Mutation

Statistical results reported in the main text are upheld if we use the average tolerance and fitness results from lines that contain the same ergosterol mutation. For all three assays conducted in nystatin (IC_{50}, growth rate, OD48), and both assays conducted in YPD (growth rate and OD48), we recover the same result previously reported, that only the ergosterol gene that bears a mutation has a significant effect on the results (i.e., neither mutation type nor their interaction).

Table C.1: IC_{50} in YPD+4μM nystatin

	F value	df	p-value
Gene	318.43	3	< 0.00001
Type of mutation	1.41	2	0.28
Interaction	0.52	2	0.61

Table C.2: Growth rate in YPD+4μM nystatin

	F value	df	p-value
Gene	11.88	3	0.0007
Type of mutation	2.98	2	0.09
Interaction	0.08	2	0.93

Table C.3: OD48 YPD+4μM nystatin

	F value	df	p-value
Gene	5.16	3	0.016
Type of mutation	0.67	2	0.53
Interaction	0.42	2	0.67

Table C.4: Growth rate in YPD

	F value	df	p-value
Gene	10.71	3	0.001
Type of mutation	0.27	2	0.77
Interaction	0.33	2	0.72

Table C.5: OD48 in YPD

	F value	df	p-value
Gene	4.64	3	0.02
Type of mutation	0.09	2	0.91
Interaction	0.73	2	0.50

The statistical results of our correlation tests between different fitness proxies also yield the same results if we average across line replicates with the same mutation. Comparing between nystatin tolerance breadth (IC_{50}) and two fitness proxies in nystatin we find that all three assays are significantly correlated to each other (growth rate and OD48 in nystatin: cor = 0.88, t_{18} = 7.9, $p < 0.0001$; growth rate and IC_{50}: cor = 0.80, t_{18} = 5.7, $p < 0.0001$; OD48 and IC_{50}: cor = 0.72, t_{18}= 4.4,p = 0.0003). When we compare IC_{50} in nystatin and the same two fitness proxies when the lines are grown in YPD we find the same result we previously reported, i.e., the only significant correlation is growth rates in YPD with IC_{50} in nystatin (growth rate and OD48 in YPD: cor = 0.35, t_{18} = 1.6, p = 0.13; growth rate in YPD and IC_{50} in nystatin: cor = 0.65, t_{18} = 3.6, p = 0.002; OD48 and IC_{50} in nystatin: cor = 0.20, t_{18} = 0.9, p = 0.4).

Finally, we also recover the same pattern of tradeoffs between tolerance to nystatin and secondary environments (all measured as IC_{50}). Specifically, we find tolerance to nystatin is significantly correlated to both ethanol (cor = -0.62, t_{19} = -3.5, p = 0.002) and copper (cor = -0.88, t_{19} = -8.0, $p < 0.0001$), but not to salt (cor = 0.13, t_{19} = 0.6, p = 0.56).

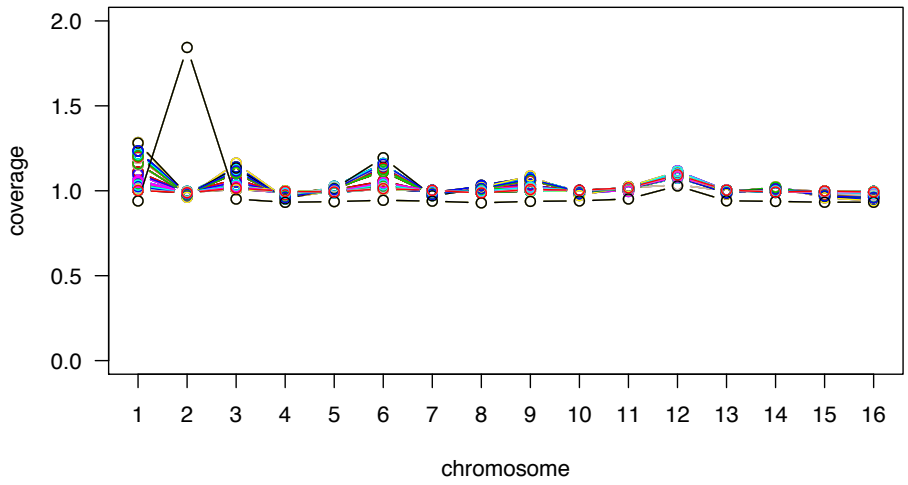


Figure C.1: Relative coverage of each chromosome from genomic alignments. Using the Illumina genomic sequence data, the total coverage for each chromosome was calculated as the proportion of sequenced sites mapping to a particular chromosome relative to the proportion of known mapped sites located on that chromosome within the yeast reference genome (as reported by configureBuild.pl in Illumina's CASAVA-1.8.0 package). Examining the coverage data for each chromosome from each BMN line (each line is plotted with a unique symbol) indicates only one aneuploidy event - an extra copy of chromosome 2 in BMN27.

Table C.6: The date mutations were acquired. Mutations were acquired in two screens ('a' & 'b'), with each acquisition experiment lasting one week. Different ancestral colonies were used to initiate each acquisition screen. By examining the date of isolation and screen we gain insight into the process that led to identical ergosterol mutations (BMN2-4, BMN7-10, BMN11-15, BMN17-20 and BMN24-27).

BMN	Date Isolated	Screen
1	10.07.25	a
2	10.08.14	b
3	10.08.14	b
4	10.08.09	b
5	10.07.23	a
6	10.07.25	a
7	10.07.25	a
8	10.07.25	a
9	10.07.25	a
10	10.07.23	b
11	10.08.09	b
12	10.08.09	b
13	10.08.09	b
14	10.08.09	b
15	10.08.09	a
16	10.07.23	a
17	10.07.25	a
18	10.07.25	a
19	10.07.23	a
20	10.07.23	b
21	10.08.10	b
22	10.08.12	b
23	10.08.12	b
24	10.08.10	b
25	10.08.12	b
26	10.08.12	b
27	10.08.13	a
28	10.07.25	a
29	10.07.25	b
30	10.08.10	b
31	10.08.10	b
32	10.08.12	b
33	10.08.10	b
34	10.08.13	b
35	10.08.14	b

Table C.7: Mutations in genes not in the ergosterol biosynthesis pathway. Sixteen additional unique mutations plus one aneuploidy were found within 17 lines.

1	YOL073C	XV.193885	916	G>A	Asp306Asn
3	YPL039W	XVI.479629	405	G>A	synonymous (Arg)
5	*CDC23*	VIII.438829	222	A>C	synonymous (Ile)
8		X.30640		A>C	
11-15	*GDA1*	V.74568	798	C>T	synonymous (Val)
15	*MBP1*	IV.354071	1195	T>G	Phe399Val
16		XIV.1753512		A>T	
16		XIV.1753521		C>A	
22	*COX1*	mt.23360	9543	T>A	synonymous (Ile)
22	*COX1*	mt.23361	9544	A>T	Ile3182Leu
23	*SCW11*	VII.442319	591	A>G	synonymous (Ser)
24-27	*FCY2*	V.267873	241	G>A	Glu81Lys
26		XIV.507563		T>G	
27	*AUR1*	XI.436609	1030	C>T	Pro344Ser
27	YJR107W	X.627995	656	G>A	Trp219Stop
27				+Chr2	
29	*SGS1*	XIII.644130	1129	A>G	Asn377Asp
30	*MDM20*	XV.188973	1950	T>G	Ile650Met
31	*ALT1*	XII.319765	251	T>C	Leu84Pro
35		I.73925		A>G	Met>Thr

Table C.8: Maximum likelihood results for growth in nystatin. We fit a likelihood model to the combined data for each mutation line and the ancestral line. The full model allowed two IC_{50} values to be fit to the data, while the constrained model forced the IC_{50} for the mutant and ancestral lines to be equal. All mutation lines were found to have significantly different IC_{50} values from the ancestor (difference in log-likelihood between the two models > 1.92; see Methods).

BMN	Gene	LogLikelihood - FullModel	LogLikelihood - Constrained	Difference in Models
	*erg6*Δ	35.60	-12.04	47.64
	*erg5*Δ	18.18	-10.64	28.82
	*erg3*Δ	46.23	5.76	40.46
1	*ERG7*	18.12	10.77	7.34
2	*ERG6*	50.14	-5.47	55.61
3	*ERG6*	38.97	-16.94	55.91
4	*ERG6*	60.91	-18.95	79.86
5	*ERG6*	37.31	-13.68	50.99
6	*ERG6*	60.87	-0.47	61.34
7	*ERG6*	53.25	-2.56	55.81
8	*ERG6*	49.38	-16.30	65.68
9	*ERG6*	35.15	-16.92	52.08
10	*ERG6*	42.37	-13.91	56.28
11	*ERG6*	62.46	6.72	55.74
12	*ERG6*	63.21	7.17	56.04
13	*ERG6*	69.14	12.87	56.27
14	*ERG6*	38.78	-16.81	55.59
15	*ERG6*	51.53	-17.35	68.89
16	*ERG6*	39.39	-15.73	55.12
17	*ERG6*	68.80	12.80	56.00
18	*ERG6*	35.34	-13.35	48.69
19	*ERG6*	35.71	-16.51	52.22
20	*ERG6*	42.61	-12.49	55.10
21	*ERG3*	43.94	8.56	35.37
22	*ERG3*	33.01	-0.17	33.18
23	*ERG3*	31.41	-0.68	32.09
24	*ERG3*	31.01	6.83	24.18
25	*ERG3*	28.70	2.33	26.37
26	*ERG3*	57.92	14.94	42.98
27	*ERG3*	60.53	23.06	37.47
28	*ERG3*	34.21	-3.92	38.13
29	*ERG3*	41.40	4.49	36.91
30	*ERG3*	48.29	17.20	31.09
31	*ERG3*	29.86	-0.50	30.36
32	*ERG3*	35.36	9.62	25.74
33	*ERG3*	29.21	-0.11	29.32
34	*ERG3*	25.92	1.46	24.46
35	*ERG5*	-3.30	-14.00	10.70

Table C.9: T-test results comparing growth rate of BMN lines in nystatin to five ancestral colonies.

BMN	t	df	p
1	-3.4	18.7	0.003
2	-15.5	35.8	< 0.0001
3	-4.1	6.9	0.005
4	-4.1	6.8	0.005
5	-7.1	8.7	< 0.0001
6	-16	33.6	< 0.0001
7	-5.9	9.3	0.0002
8	-12.4	20.3	< 0.0001
9	-7.6	14.6	< 0.0001
10	-4.1	6.1	0.006
11	-16.8	33.6	< 0.0001
12	-7.6	8.6	< 0.0001
13	-9.1	7.1	< 0.0001
14	-15.8	30.8	< 0.0001
15	-6.6	7.5	0.0002
16	-2.4	3.3	0.087
17	-5.4	8.4	0.0005
18	-4	9.2	0.003
19	-6.7	7.6	0.0002
20	-17	32.5	< 0.0001
21	-1.7	3.4	0.18
22	-3.1	5.3	0.026
23	-2.4	5.2	0.06
24	-4.4	7.8	0.002
25	-3.1	10.3	0.011
26	-6.2	23	< 0.0001
27	-2.6	6.8	0.034
28	-3.5	3.7	0.027
29	-2.5	6	0.046
30	-4.1	6.5	0.0056
31	-2.9	5.5	0.029
32	-2.9	7.2	0.021
33	-5.1	33.8	< 0.0001
34	-5.4	7.2	0.0009
35	-2.1	4.4	0.10

Table C.10: T-test results comparing OD48 of BMN lines grown in nystatin to five ancestral colonies.

BMN	t	df	p
1	-5.5	8.2	0.0005
2	-10.8	5.2	0.0001
3	-8.1	5.1	0.0004
4	-7.1	5.1	0.0008
5	-10.5	5.1	0.0001
6	-12.1	3.1	0.001
7	-9.2	6.2	0.0001
8	-13.7	3.1	0.0007
9	-8.7	8.2	$<$ 0.0001
10	-4.5	5	0.0066
11	-9.8	3.1	0.002
12	-10.9	5.2	0.0001
13	-14	3.2	0.0006
14	-6.4	2	0.023
15	-8.4	5.1	0.0004
16	-2.6	3	0.082
17	-5.9	6.1	0.001
18	-5.3	7.1	0.001
19	-6.1	5.1	0.002
20	-11.3	4.1	0.0003
21	-2.8	3	0.065
22	-4.4	4.1	0.012
23	-3.8	4	0.018
24	-7.2	5.2	0.0007
25	-8.2	6.2	0.0001
26	-8.8	4.3	0.0007
27	-4.4	5.1	0.007
28	-3.8	3	0.032
29	-3	5.1	0.030
30	-6	5.1	0.002
31	-10.8	4.4	0.0003
32	-7	5.2	0.0007
33	-6.1	3.1	0.008
34	-8.8	5.1	0.0003
35	-2.3	4	0.087

Table C.11: T-test results comparing growth rate of BMN lines grown in YPD to five ancestral colonies.

BMN	t	df	p
1	6	20.6	< 0.0001
2	4	20.1	0.0008
3	3.7	13.7	0.0026
4	4.7	21	0.0001
5	5	20.4	0.0001
6	3.3	21.6	0.0033
7	3.5	21.8	0.0019
8	3.2	7	0.0146
9	3.9	21.8	0.0007
10	5.8	21.8	< 0.0001
11	3.9	5.2	0.011
12	4.7	18	0.0002
13	3.7	9.1	0.0045
14	4.1	19.9	0.0006
15	4.9	20.4	0.0001
16	4.9	10.1	0.0006
17	3.9	17	0.0012
18	4.4	22	0.0002
19	5	20.1	0.0001
20	5.5	19.4	< 0.0001
21	6.8	7.2	0.0002
22	14.3	21.8	< 0.0001
23	14	22	< 0.0001
24	6.8	9.3	0.0001
25	10.9	12.6	< 0.0001
26	13.4	21.3	< 0.0001
27	5.6	11.1	0.0002
28	10.1	12.9	< 0.0001
29	5.1	8.8	0.0007
30	7.6	14.1	< 0.0001
31	5.1	6.6	0.0017
32	12.6	21.6	< 0.0001
33	4.5	6.7	0.0029
34	7.5	7.4	0.0001
35	2.6	11.4	0.022

Table C.12: T-test results comparing OD48 of BMN lines grown in YPD to five ancestral colonies.

BMN	t	df	p
1	5.5	12	0.0001
2	6	19	<0.0001
3	3.8	4	0.020
4	12.2	7.9	< 0.0001
5	3.5	4	0.024
6	4.5	4	0.011
7	10.1	5.9	0.0001
8	7.3	7.7	0.0001
9	12.3	9.5	< 0.0001
10	5.3	4.9	0.003
11	5.2	5.9	0.0022
12	4.7	4.9	0.0058
13	7	8.2	0.0001
14	3.1	3.4	0.043
15	8.8	11.5	< 0.0001
16	13.7	21.8	< 0.0001
17	3.6	3.7	0.026
18	2.8	3.3	0.063
19	4.4	3.4	0.016
20	3.5	3.8	0.027
21	10.1	5.7	0.0001
22	17.8	21.7	< 0.0001
23	4.3	3.7	0.015
24	6.2	5.9	0.0009
25	8	5.5	0.0003
26	9.2	8.6	< 0.0001
27	10	6	0.0001
28	10.6	9.5	< 0.0001
29	7.8	3.9	0.0016
30	7.7	5	0.0006
31	18.4	20.4	< 0.0001
32	5	4.5	0.0052
33	15.7	22	< 0.0001
34	5.7	7.6	0.0006
35	2.3	4.1	0.079

Table C.13: Likelihood ratio tests comparing IC_{50} of ancestral and BMN lines in copper, ethanol and salt.

		Copper ($CuSO_4$)	Ethanol	Salt (NaCl)
BMN	Gene	Difference in Models	Difference in Models	Difference in Models
1	*ERG7*	3.76	30.00	4.79
2	*ERG6*	5.30	11.75	23.87
3	*ERG6*	0.62	5.18	23.50
4	*ERG6*	2.15	12.06	27.00
5	*ERG6*	1.73	24.35	26.36
6	*ERG6*	12.24	25.31	19.81
7	*ERG6*	1.93	14.42	20.05
8	*ERG6*	2.58	17.05	26.35
9	*ERG6*	2.61	17.76	30.26
10	*ERG6*	3.05	15.25	34.54
11	*ERG6*	2.99	8.05	18.08
12	*ERG6*	2.39	4.19	25.83
13	*ERG6*	1.88	18.78	32.90
14	*ERG6*	13.63	17.55	27.48
15	*ERG6*	1.91	29.10	29.00
16	*ERG6*	4.34	20.17	26.91
17	*ERG6*	0.44	6.95	23.83
18	*ERG6*	3.49	21.75	27.88
19	*ERG6*	1.21	20.40	24.13
20	*ERG6*	1.94	33.79	28.62
21	*ERG3*	2.20	0.47	48.55
22	*ERG3*	2.00	0.02	49.20
23	*ERG3*	2.50	1.00	48.42
24	*ERG3*	10.06	0.02	9.90
25	*ERG3*	2.17	0.15	40.52
26	*ERG3*	5.49	0.48	40.25
27	*ERG3*	9.99	0.49	0.44
28	*ERG3*	5.48	0.58	45.23
29	*ERG3*	9.82	< 0.0001	45.53
30	*ERG3*	1.18	0.01	44.65
31	*ERG3*	3.20	0.15	40.96
32	*ERG3*	2.48	< 0.0001	47.08
33	*ERG3*	1.19	0.11	49.30
34	*ERG3*	1.89	0.08	8.45
35	*ERG5*	2.56	0.33	0.99

Appendix D

Appendix for Chapter 5: Mutational effects depend on ploidy level: All else is not equal

D.1 Supporting Tables

Table D.1: Linear mixed-effects models that account for batch effects in the growth assays show similar statistical results as a two-way ANOVA (presented in the main text).

	Linear mixed-effects model		
Environment	ploidy	gene	interaction
2μM nystatin	$F_{1,71} = 9.32$, $p < 0.0001$	$F_{3,71} = 121.46$, $p < 0.0001$	$F_{3,71} = 0.49$, $p = 0.69$
4μM nystatin	$F_{1,63} = 6.58$, $p = 0.013$	$F_{3,63} = 30.39$, $p < 0.0001$	$F_{3,63} = 0.36$, $p = 0.78$

Table D.2: Line-specific statistical results for dose-response assay parameters between haploid and homozygous diploids. Parameters are significant between ploidy levels if the difference in log-likelihood between full and constrained models is greater than 1.92 (see Methods, significance indicated with *).

	LogLik full – constrained model		
BMN Line	Tolerance (IC_{50})	asymptote (a)	slope (m_{50})
1	7.13*	< 0.01	2.83*
3	0.04	8.62*	2.59*
5	59.44*	13.96*	0.44
6	35.57*	32.38*	< 0.01
9	8.94*	21.36*	0.67
13	42.13*	12.66*	< 0.01
16	0.15	7.85*	0.05
19	39.11*	26.3*	< 0.01
21	16.25*	0.14	4.01*
22	8.01*	0.31	12.17*
23	12.18*	1.21	5.62*
25	6.34*	0.04	4.52*
28	16.19*	< 0.01	0.02
29	4.14*	1.45	0.81
30	5.03*	0.28	8.20*
31	2.80*	0.19	0.48
33	6.61*	0.04	0.66
32	0.59	0.38	3.25*
34	35.23*	0.12	<0.01
35	0.06	0.10	< 0.01

Table D.3: Line-specific statistical results growth rate assays between haploid and homozygous diploid backgrounds. Growth rate significance determined with t-tests between haploids and diploids; * denotes $p < 0.05$, $^{+}$ denotes $p < 0.10$.

	Growth rate		
BMN Line	2μM nystatin	4μM nystatin	8μM nystatin
1	$t_{9.4}$=3.1 p = 0.012*	$t_{14.0}$=1.5, p = 0.16	$t_{7.2}$=2.0, p = 0.078^{+}
3	$t_{9.7}$=3.7, p = 0.004*	$t_{9.4}$=6.1, p = 0.0001*	$t_{14.0}$=2.2, p = 0.045*
5	$t_{12.1}$=1.7, p = 0.11	$t_{12.9}$=1.3, p = 0.23	$t_{13.7}$=4.3, p = 0.0008*
6	$t_{13.9}$=6.8, $p < 0.0001$*	$t_{14.1}$=2.8, p = 0.014*	$t_{12.0}$=2.3, p = 0.037*
9	$t_{8.9}$=2.2, p = 0.052^{+}	$t_{13.5}$=-2.2, p = 0.045*	$t_{13.7}$=2.4, p = 0.032*
13	$t_{14.0}$=1.8, p = 0.095^{+}	$t_{5.1}$=0.5, p = 0.63	$t_{14.0}$=1.9, p = 0.077^{+}
16	$t_{8.1}$=2.3, p = 0.051^{+}	$t_{6.0}$=2.0, p = 0.090^{+}	$t_{13.3}$=11.6, $p < 0.0001$*
19	$t_{13.8}$=5.1, p = 0.0001*	$t_{10.6}$=1.4, p = 0.19	$t_{12.2}$=2.9, p = 0.012*
21	$t_{10.8}$=1.4, p = 0.18	$t_{11.0}$=0.9, p = 0.45	$t_{8.1}$=0.7, p = 0.48
22	$t_{10.7}$=1.2, p = 0.26	$t_{12.1}$=0.2, p = 0.82	$t_{7.9}$=1.2, p = 0.28
23	$t_{12.1}$=0.6, p = 0.56	$t_{11.9}$=1.0, p = 0.32	$t_{7.1}$=1.4, p = 0.20
25	$t_{13.6}$=2.2, p = 0.049*	$t_{12.2}$=1.5, p = 0.17	$t_{7.1}$=1.5, p = 0.17
28	$t_{13.6}$=3.0, p = 0.011*	$t_{3.8}$=22.5, $p < 0.0001$*	$t_{7.0}$=3.1, p = 0.017*
29	$t_{10.3}$=2.4, p = 0.034*	$t_{9.2}$=0.1, p = 0.92	$t_{13.8}$=-0.2, p = 0.84
30	$t_{13.4}$=2.0, p = 0.064^{+}	$t_{14.8}$=1.9, p = 0.079^{+}	$t_{10.0}$=1.4, p = 0.19
31	$t_{8.5}$=0.7, p = 0.53	$t_{9.6}$=0.6, p = 0.55	$t_{12.2}$=-0.5, p = 0.65
33	$t_{13.7}$=1.6, p = 0.13	$t_{10.7}$=0.7, p = 0.47	$t_{7.0}$=1.6, p = 0.16
32	$t_{13.9}$=6.2, $p < 0.0001$*	$t_{13.0}$=1.2, p = 0.25	$t_{7.0}$=1.8, p = 0.11
34	$t_{7.5}$=-7.1, p = 0.0002*	$t_{11.4}$=-1.9, p = 0.085^{+}	$t_{10.3}$=-0.5, p = 0.63
35	$t_{9.4}$=0.5, p = 0.60	$t_{4.3}$=-0.4, p = 0.70	$t_{14.0}$=0.04, p = 0.97

D.2 Supporting Figures

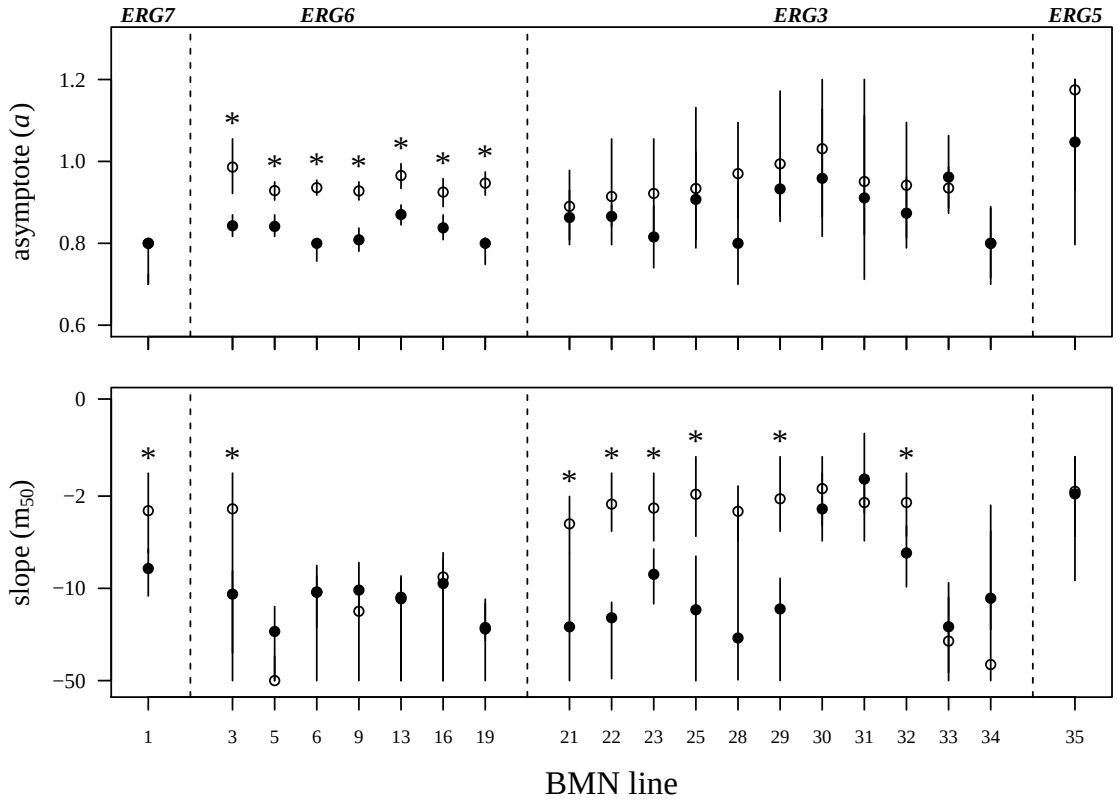


Figure D.1: Nystatin beneficial mutation lines tend to have a higher asymptote at low levels of nystatin (top graph) and higher slope at IC_{50} (bottom) in haploids compared to homozygous diploids.

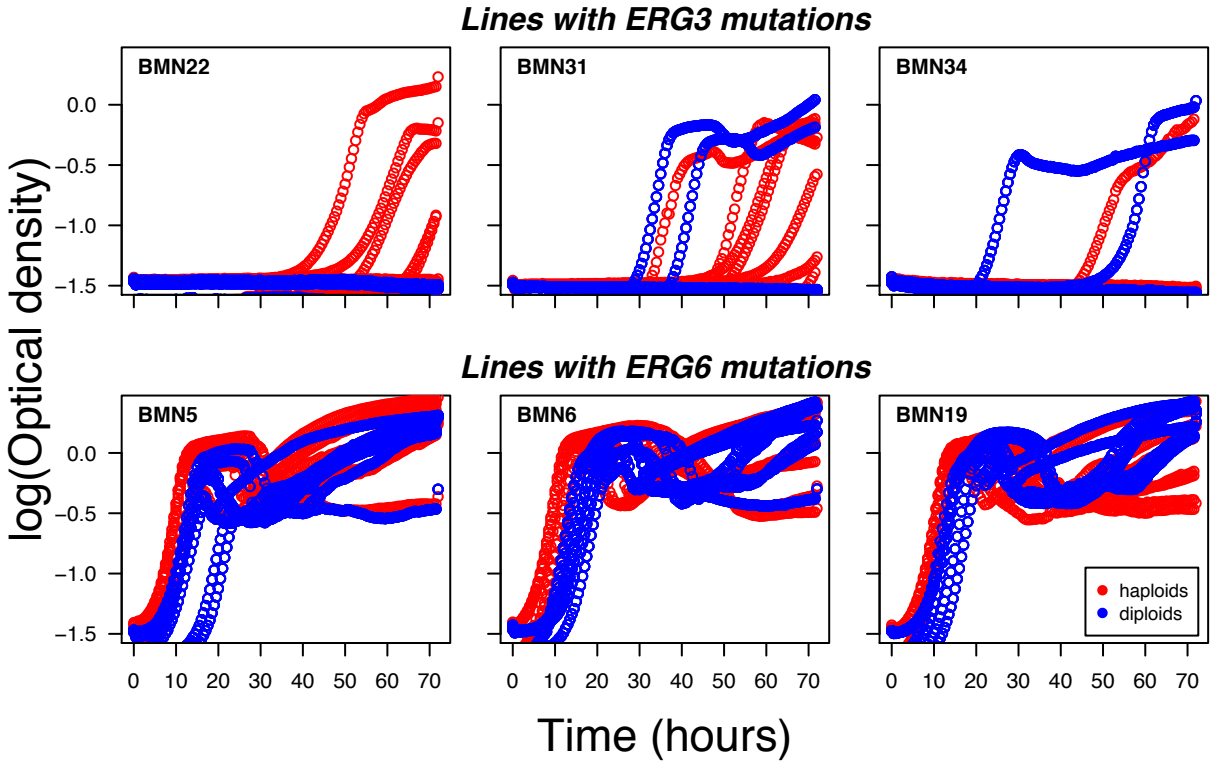


Figure D.2: Lines with mutations in *ERG3* grow stochastically in YPD+8*mu*M nystatin (top three panels), while lines with mutations in *ERG6* continue to grow fairly consistently (bottom three panels). Variation in optical density following the rapid phase of growth likely reflects variation between wells in cell clumping and settling. Note that we use the intrinsic growth rate during rapid growth as a fitness proxy.

Appendix E

Appendix for Chapter 6: Unstable heterozygotes

E.1 Supporting Tables

Table E.1: All homozygous mutation lines grew slower and reached lower biomass than wildtype in an unstressful environment(YPD).

BMN Line	Gene	Growth rate	Biomass production
1	*ERG7*	$t_{31.5} = 14.36$, $p < 0.0001$	$t_{29.4} = 6.00$, $p < 0.0001$
3	*ERG6*	$t_{31.5} = 17.40$, $p < 0.0001$	$t_{15.4} = 9.49$, $p < 0.0001$
5	*ERG6*	$t_{30.0} = 13.56$, $p < 0.0001$	$t_{14.3} = 6.90$, $p < 0.0001$
6	*ERG6*	$t_{35.7} = 13.40$, $p < 0.0001$	$t_{26.8} = 5.97$, $p < 0.0001$
9	*ERG6*	$t_{34.5} = 14.00$, $p < 0.0001$	$t_{31.5} = 5.77$, $p < 0.0001$
13	*ERG6*	$t_{38.0} = 12.48$, $p < 0.0001$	$t_{21.4} = 14.32$, $p < 0.0001$
16	*ERG6*	$t_{38.0} = 14.22$, $p < 0.0001$	$t_{10.8} = 8.08$, $p < 0.0001$
20	*ERG6*	$t_{38.0} = 16.04$, $p < 0.0001$	$t_{11.2} = 6.83$, $p < 0.0001$
21	*ERG3*	$t_{37.5} = 13.94$, $p < 0.0001$	$t_{24.7} = 15.24$, $p < 0.0001$
22	*ERG3*	$t_{31.3} = 17.49$, $p < 0.0001$	$t_{18.3} = 13.32$, $p < 0.0001$
23	*ERG3*	$t_{32.7} = 17.05$, $p < 0.0001$	$t_{18.9} = 14.24$, $p < 0.0001$
25	*ERG3*	$t_{29.8} = 17.14$, $p < 0.0001$	$t_{18.2} = 14.63$, $p < 0.0001$
28	*ERG3*	$t_{34.8} = 16.90$, $p < 0.0001$	$t_{13.5} = 9.41$, $p < 0.0001$
29	*ERG3*	$t_{37.0} = 12.62$, $p < 0.0001$	$t_{30.9} = 18.34$, $p < 0.0001$
30	*ERG3*	$t_{37.9} = 13.77$, $p < 0.0001$	$t_{34.3} = 21.30$, $p < 0.0001$
31	*ERG3*	$t_{37.2} = 14.44$, $p < 0.0001$	$t_{26.9} = 15.44$, $p < 0.0001$
32	*ERG3*	$t_{31.5} = 17.29$, $p < 0.0001$	$t_{34.1} = 22.10$, $p < 0.0001$
33	*ERG3*	$t_{36.4} = 14.96$, $p < 0.0001$	$t_{31.5} = 18.54$, $p < 0.0001$
34	*ERG3*	$t_{34.1} = 14.86$, $p < 0.0001$	$t_{37.7} = 12.06$, $p < 0.0001$
35	*ERG5*	$t_{32.0} = 5.39$, $p < 0.0001$	$t_{19.4} = 5.27$, $p < 0.0001$

Table E.2: Heterozygous maximal growth rates compared to wildtype and homozygous mutant diploids. Dominance of mutations for maximal growth rate is inferred from the statistical results.

BMN Line	Gene	Compared to wildtype diploid	Compared to homozygous mutant	Significance
1	*ERG7*	$t_{10.9}$ = -1.14, p = 0.28	$t_{11.4}$ = 2.26, p = 0.044	R
3	*ERG6*	$t_{33.5}$ = 13.34, $p < 0.0001$	$t_{12.2}$ = 1.51, p = 0.16	D
5	*ERG6*	$t_{11.2}$ = -0.40, p = 0.69	$t_{16.0}$ = 3.72, p = 0.0019	O
6	*ERG6*	$t_{32.5}$ = 12.04, $p < 0.0001$	$t_{9.1}$ = 1.39, p = 0.20	D
9	*ERG6*	$t_{25.5}$ = -3.36, p = 0.002	$t_{12.2}$ = 3.41, p = 0.005	I
13	*ERG6*	$_{11.6}t$ = 0.13, p = 0.90	$t_{12.7}$ = 6.08, $p < 0.0001$	R
16	*ERG6*	$t_{14.0}$ = -1.53, p = 0.15	$t_{16.5}$ = 6.10, $p < 0.0001$	R
20	*ERG6*	$t_{12.2}$ = 1.18, p = 0.26	$t_{17.8}$ = 4.07, p = 0.0007	R
21	*ERG3*	$t_{15.3}$ = -1.57, p 0.14	$t_{13.2}$ = 6.80, $p < 0.0001$	R
22	*ERG3*	$t_{37.8}$ = 13.16, $p < 0.0001$	$t_{10.9}$ = 9.33, $p < 0.0001$	I
23	*ERG3*	$t_{17.0}$ = 5.38, $p < 0.0001$	$t_{10.2}$ = 11.57, $p < 0.0001$	I
25	*ERG3*	$t_{13.8}$ = 2.30, p = 0.037	$t_{15.9}$ = 7.85, $p < 0.0001$	I
28	*ERG3*	$t_{11.7}$ = -1.97, p = 0.072	$t_{17.7}$ = 5.95, $p < 0.0001$	R
29	*ERG3*	$t_{11.4}$ = -1.61, p = 0.13	$t_{11.4}$ = 9.00, $p < 0.0001$	R
30	*ERG3*	$t_{14.7}$ = -0.99, p = 0.34	$t_{10.5}$ = 6.89, $p < 0.0001$	R
31	*ERG3*	$t_{11.6}$ = -0.49, p = 0.63	$t_{12.8}$ = 7.00, $p < 0.0001$	R
32	*ERG3*	$t_{11.7}$ = 2.31, p = 0.040	$t_{11.0}$ = 10.83, $p < 0.0001$	I
33	*ERG3*	$t_{11.6}$ = -1.69, p = 0.12	$t_{13.0}$ = 12.48, $p < 0.0001$	R
34	*ERG3*	$t_{33.3}$ = 13.16, $p < 0.0001$	$t_{13.2}$ = 9.19, $p < 0.0001$	I
35	*ERG5*	$t_{9.04}$ = -0.88, p = 0.40	$t_{9.9}$ = -0.58, p = 0.57	U

R= recessive (heterozygote is not significantly different than wildtype)
D = dominant (heterozygote is not significantly different than homozygous mutant)
I = intermediate (heterozygote is significantly different than both)
O = overdominant (heterozygote grows significantly better than wildtype)
U = unclear (heterozygote is not significantly different than either)

Table E.3: Heterozygous maximal growth rates compared to wildtype and homozygous mutant diploids. Dominance of mutations for biomass production is inferred from the statistical results.

BMN Line	Gene	Compared to wildtype diploid	Compared to homozygous mutant	Significance
1	*ERG7*	$t_{11.9} = 0.77$, $p = 0.46$	$t_{11.4} = 2.26$, $p = 0.044$	R
3	*ERG6*	$t_{33.3} = 8.04$, $p < 0.0001$	$t_{12.2} = 1.51$, $p = 0.16$	D
5	*ERG6*	$t_{11.4} = 0.81$, $p = 0.43$	$t_{16.0} = 3.72$, $p = 0.0019$	R
6	*ERG6*	$t_{34.0} = 6.20$, $p < 0.0001$	$t_{9.1} = 1.39$, $p = 0.20$	D
9	*ERG6*	$t_{13.7} = 0.06$, $p = 0.95$	$t_{12.2} = 3.41$, $p = 0.0050$	R
13	*ERG6*	$t_{11.6} = 1.71$, $p = 0.11$	$t_{12.7} = 6.08$, $p < 0.0001$	R
16	*ERG6*	$t_{12.4} = 1.14$, $p = 0.28$	$t_{16.5} = 6.10$, $p < 0.0001$	R
20	*ERG6*	$t_{11.7}$ =1.82, $p = 0.09$	$t_{17.8} = 4.07$, $p = 0.0007$	R
21	*ERG3*	$t_{12.8} = 2.37$, $p = 0.035$	$t_{13.2} = 6.80$, $p < 0.0001$	I
22	*ERG3*	$t_{37.4} = 8.33$, $p < 0.0001$	$t_{10.9} = 9.33$, $p < 0.0001$	I
23	*ERG3*	$t_{35.2} = 7.30$, $p < 0.0001$	$t_{10.2} = 11.60$, $p < 0.0001$	I
25	*ERG3*	$t_{13.1} = 2.32$, $p = 0.037$	$t_{15.9} = 7.85$, $p < 0.0001$	I
28	*ERG3*	$t_{12.5} = 1.20$, $p = 0.25$	$t_{17.7} = 5.95$, $p < 0.0001$	R
29	*ERG3*	$t_{12.2} = 0.55$, $p = 0.59$	$t_{11.4} = 8.97$, $p < 0.0001$	R
30	*ERG3*	$t_{11.5} = 2.58$, $p = 0.025$	$t_{10.5} = 6.89$, $p < 0.0001$	I
31	*ERG3*	$t_{13.1} = 2.30$, $p = 0.039$	$t_{12.8} = 6.97$, $p < 0.0001$	I
32	*ERG3*	$t_{12.5} = 0.64$, $p = 0.53$	$t_{11.04} = 10.82$, $p < 0.0001$	R
33	*ERG3*	$t_{15.0}$ = -0.19, $p = 0.86$	$t_{13.0} = 12.48$, $p < 0.0001$	R
34	*ERG3*	$t_{32.3} = 8.13$, $p < 0.0001$	$t_{13.2} = 9.19$, $p < 0.0001$	I
35	*ERG5*	$t_{8.9} = 1.32$, $p = 0.22$	$t_{9.9}$ = -0.58, $p = 0.57$	U

R= recessive (heterozygote is not significantly different than wildtype)
D = dominant (heterozygote is not significantly different than homozygous mutant)
I = intermediate (heterozygote is significantly different than both)
O = overdominant (heterozygote grows significantly better than wildtype)
U = unclear (heterozygote is not significantly different than either)

Table E.4: Heterozygous colonies isolated and Sanger sequenced after growth in YPD in bioscreen wells (b) or deep well plates (d).

BMN	Gene	Genotype at initial locus	Growth Experiment	Colony
1	*ERG7*	het	d	7A
3	*ERG5*	het	d	9B
6	*ERG6*	het	d	8C
9	*ERG6*	het	d	5C
13	*ERG6*	het	d	10A
20	*ERG6*	het	d	2C
21	*ERG3*	het	b	13
21	*ERG3*	het	b	93
23	*ERG3*	het	b	14
23	*ERG3*	het	b	94
32	*ERG3*	het	b	17
32	*ERG3*	het	b	97
35	*ERG5*	het	d	10B

Table E.5: Heterozygous colonies isolated and Sanger sequenced after growth in YPD+1μM nystatin in deep well plates.

BMN	Gene	Genotype at initial locus	Growth Experiment	Colonyt
1	*ERG7*	hom	d	7E
1	*ERG7*	hom	d	7G
3	*ERG6*	hom	d	9B
5	*ERG6*	hom	d	3A
5	*ERG6*	hom	d	3C
9	*ERG6*	hom	d	5A
9	*ERG6*	hom	d	5C
13	*ERG6*	hom	d	10A
13	*ERG6*	het	d	10E
20	*ERG6*	hom	d	2A
20	*ERG6*	hom	d	2C
21	*ERG3*	hom	d	4F
21	*ERG3*	hom	d	4H
22	*ERG3*	hom	d	6H
23	*ERG3*	hom	d	5B
23	*ERG3*	hom	d	5D
28	*ERG3*	hom	d	6A
28	*ERG3*	hom	d	6C
30	*ERG3*	hom	d	11A
30	*ERG3*	hom	d	11C
31	*ERG3*	hom	d	2B
31	*ERG3*	hom	d	2D
32	*ERG3*	hom	d	8B
32	*ERG3*	hom	d	8D
35	*ERG5*	het*	d	10B
35	*ERG5*	het*	d	10D
35	*ERG5*	het*	d	10F
35	*ERG5*	het*	d	10H

*The initial allele frequency in these populations is greater than 0.5, an indication that homozygous individuals are present.

Table E.6: Heterozygous colonies isolated and Sanger sequenced after growth in YPD+2μM nystatin in bioscreen wells.

BMN	Gene	Genotype at initial locus	Growth Experiment	Colony
1	*ERG7*	hom	b	52
1	*ERG7*	hom	b	134
3	*ERG6*	wt	b	116
3	*ERG6*	hom	b	188
5	*ERG6*	hom	b	80
9	*ERG6*	hom	b	12
9	*ERG6*	hom	b	130
13	*ERG6*	hom	b	54
13	*ERG6*	hom	b	136
21	*ERG3*	hom	b	60
21	*ERG3*	hom	b	106
23	*ERG3*	wt	b	26
23	*ERG3*	hom	b	144
28	*ERG3*	hom	b	14
28	*ERG3*	wt	b	50
32	*ERG3*	hom	b	68
32	*ERG3*	wt	b	186
35	*ERG5*	wt	b	190

Table E.7: Heterozygous colonies isolated and Sanger sequenced after growth in YPD+4μM nystatin in bioscreen wells (b) or deep well plates (d).

BMN	Gene	Genotype at initial locus	Growth Experiment	Colony
1	*ERG7*	wt	b	16
1	*ERG7*	hom	b	52
1	*ERG7*	hom	d	7A
1	*ERG7*	hom	d	7E
3	*ERG6*	hom	d	9H
3	*ERG6*	hom	d	9F
5	*ERG6*	hom	d	3C
9	*ERG6*	hom	b	84
9	*ERG6*	hom	d	5A
9	*ERG6*	hom	d	5C
21	*ERG3*	hom	d	4F
21	*ERG3*	wt	d	4H
22	*ERG3*	wt	d	6B
22	*ERG3*	hom	d	6H
25	*ERG3*	hom	b	30
28	*ERG3*	hom	d	6A
28	*ERG3*	hom	d	6C
31	*ERG3*	hom	d	2B
31	*ERG3*	hom	d	2D

Table E.8: Eight het-grow lines were wildtype homozygous for the initial mutation and carried secondary homozygous mutations in either *ERG6* or *ERG3*.

			Initial mutation	Alternative ergosterol mutation	
Line	Colony (Experiment)	Environment	Gene (location in gene)	Gene (location in gene)	Amino acid change
BMN1	16 (b)	4μM nystatin	*ERG7* (bp2096)	*ERG6* (bp669)	Tyr223Stop (same as BMN19)
BMN3	116 (b)	2μM nystatin	*ERG6* (bp131)	*ERG3* (bp898)	Gly300Arg (same as BMN32)
BMN21	4H (d)	4μM nystatin	*ERG3* (bp187)	*ERG6* (bp220)	Tyr74Stop (same as BMN5)
BMN22	6H (d)	4μM nystatin	*ERG3* (bp227)	tbd	tbd
BMN23	26 (b)	2μM nystatin	*ERG3* (bp284)	*ERG3* (bp187)	Arg63Stop (same as BMN21)
BMN28	50 (b)	2μM nystatin	*ERG3* (bp640)	*ERG6* (bp279)	Gly127Arg (same as BMN9)
BMN32	186 (b)	2μM nystatin	*ERG3* (bp898)	*ERG3* (bp615)	Trp205Stop same as BMN31)
BMN35	190 (b)	2μM nystatin	*ERG5* (bp252)	*ERG6* (bp131)	Gln44Stop (same as BMN3)

Table E.9: Secondary mutations remained heterozygous in the relevant het-grow lines isolated from bioscreen experiments.

BMN	Colony	Ergosterol Gene	Environment (*mu*M nystatin)	Secondary Gene	Genome Position (Chr.Bp)	Mutation
1	52	*ERG7*	2	YOLO73C	XV.193885	C>T
1	134	*ERG7*	2	YOLO73C	XV.193885	C>T
1	52	*ERG7*	4	YOLO73C	XV.193885	C>T
5	80	*ERG6*	2	*CDC23*	VIII.438829	A>C
13	54	*ERG6*	2	*GDA1*	V.74569	C>T
13	136	*ERG6*	2	*GDA1*	V.74569	C>T
23	144	*ERG3*	2	*SCW11*	VII.442319	A>G
25	30	*ERG3*	4	*FCY2*	V.267874	G>A

Table E.10: Secondary mutations remained heterozygous in the relevant het-grow lines isolated from deep well box experiments.

BMN	Colony	Ergosterol Gene	Environment (*mu*M nystatin)	Secondary Gene	Genome Position (Chr.Bp)	Mutation
1	7E	*ERG7*	1	YOL073C	XV.193885	C>T
1	7G	*ERG7*	1	YOL073C	XV.193885	C>T
1	7A	*ERG7*	4	YOL073C	XV.193885	C>T
1	7E	*ERG7*	4	YOL073C	XV.193885	C>T
3	9B	*ERG6*	1	YPL039W	XVI.479630	G>A
3	9H	*ERG6*	4	YPL039W	XVI.479630	G>A
3	9F	*ERG6*	4	YPL039W	XVI.479630	G>A
5	3A	*ERG6*	1	*CDC23*	VIII.438829	A>C
5	3C	*ERG6*	1	*CDC23*	VIII.438829	A>C
5	3C	*ERG6*	4	*CDC23*	VIII.438829	A>C
13	A10	*ERG6*	1	*GDA1*	V.74569	C>T
13	E10	*ERG6*	1	*GDA1*	V.74569	C>T
23	5B	*ERG3*	1	*SCW11*	VII.441872	A>G
23	5D	*ERG3*	1	*SCW11*	VII.441872	A>G
30	11A	*ERG3*	1	*MDM20*	XV.188974	T>G
30	11C	*ERG3*	1	*MDM20*	XV.188974	T>G

E.2 Supporting Figures

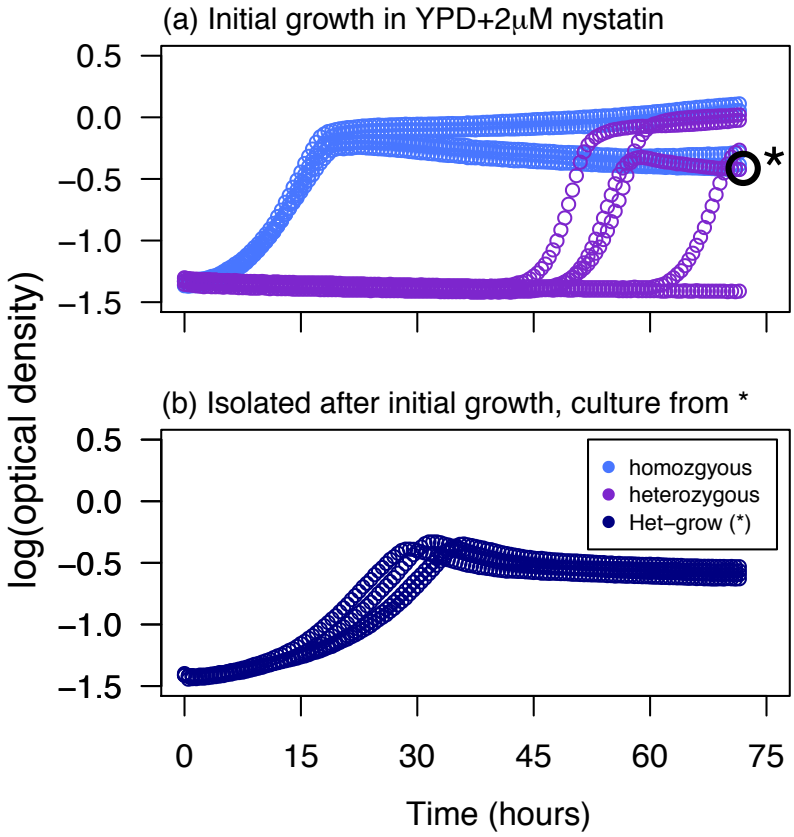


Figure E.1: Culture isolated from heterozygous replicates that (a) showed initially grew in a stressful environment grew much less stochastically after 24 hours of growth in YPD followed by inoculation back into the stressful environment (b).

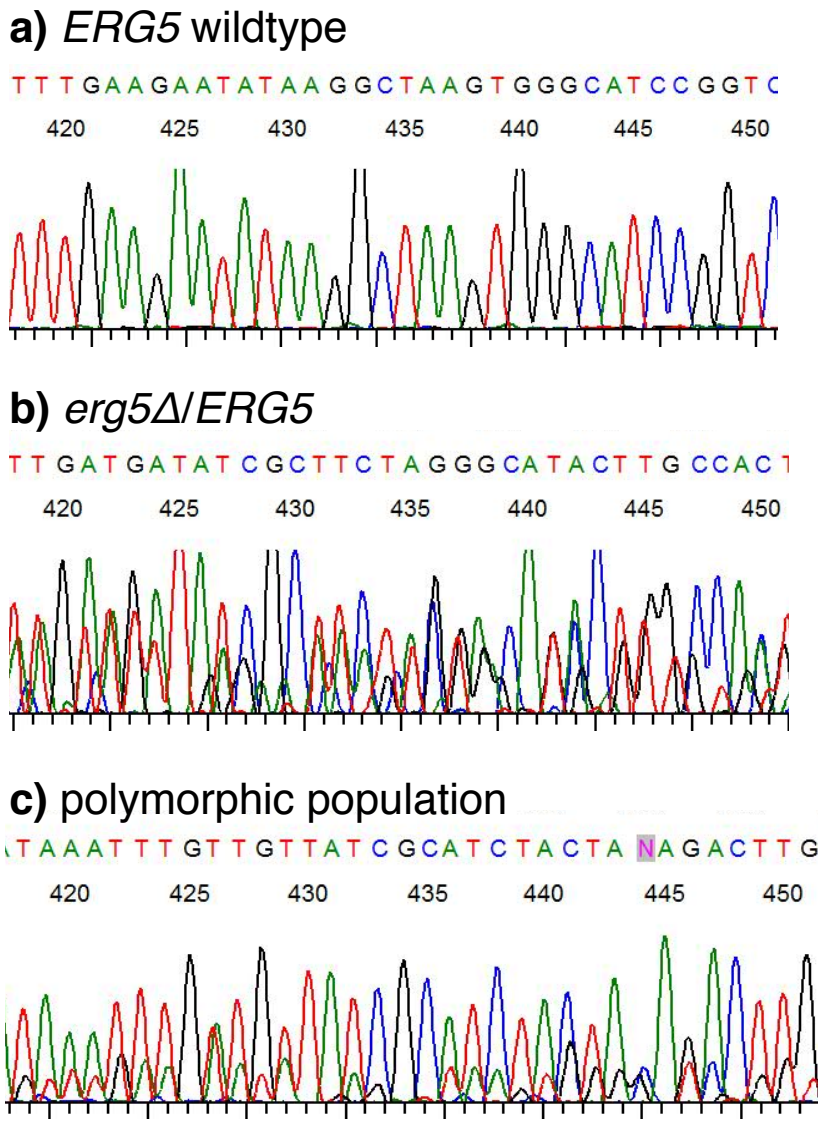


Figure E.2: Chromatograms depicting the polymorphic populations isolated from BMN35 replicates grown in YPD+1μM nystatin.

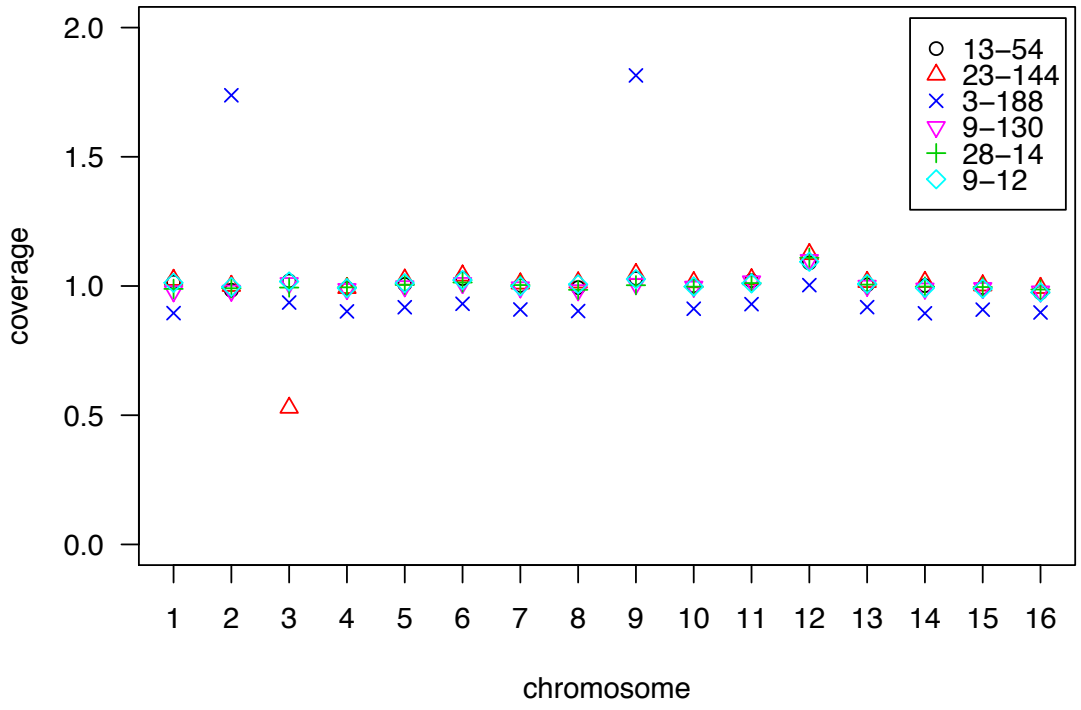
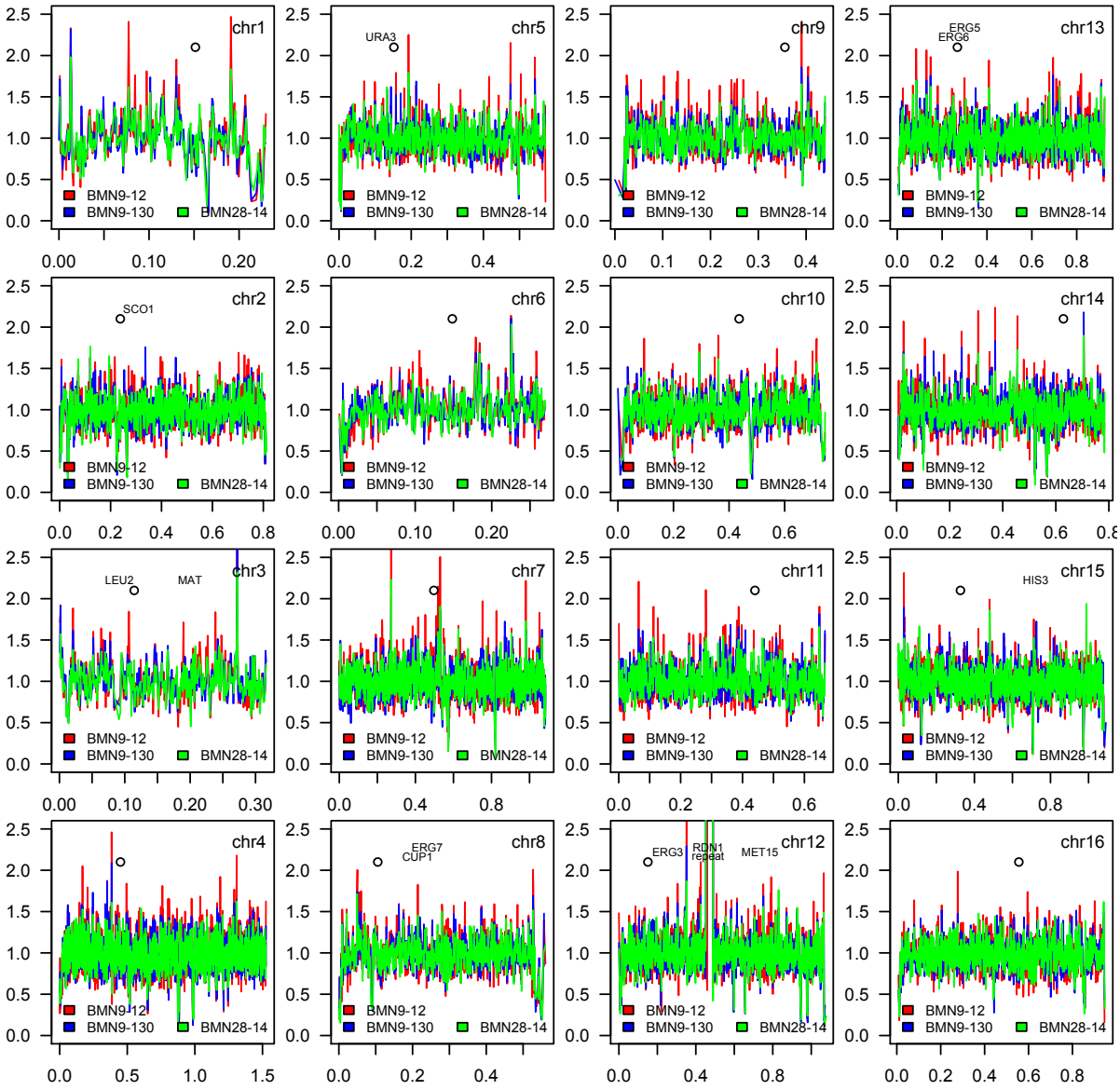


Figure E.3: For the six lines that were Illumina resequenced, coverage per chromosome was obtained using configureBuild.pl from Illumina's CASAVA-1.8.0 package. Plotted for each strain is the proportion of sequenced sites that map to each chromosome relative to the proportion of known mapped sites on that chromosome within the reference genome. Line BMN23-144 is aneuploid for chromosomes 3 (diploid according to FACS analysis, but containing only one copy of this chromosome). Line BMN3-188 was inferred to be haploid by FACS analysis, but chromosomal coverage indicates that this line retains two copies of chromosomes 2 and 9.

Panel A


chr1
chr5
URA3
chr9
chr13
ERG5
ERG6
chr2
SCO1
chr6
chr10
chr14
chr3
LEU2
MAT
chr7
chr11
chr15
HIS3
chr4
chr8
ERG7
CUP1
chr12
ERG3
RDN1
repeat
MET15
chr16
BMN9-12
BMN9-130
BMN28-14

Panel B

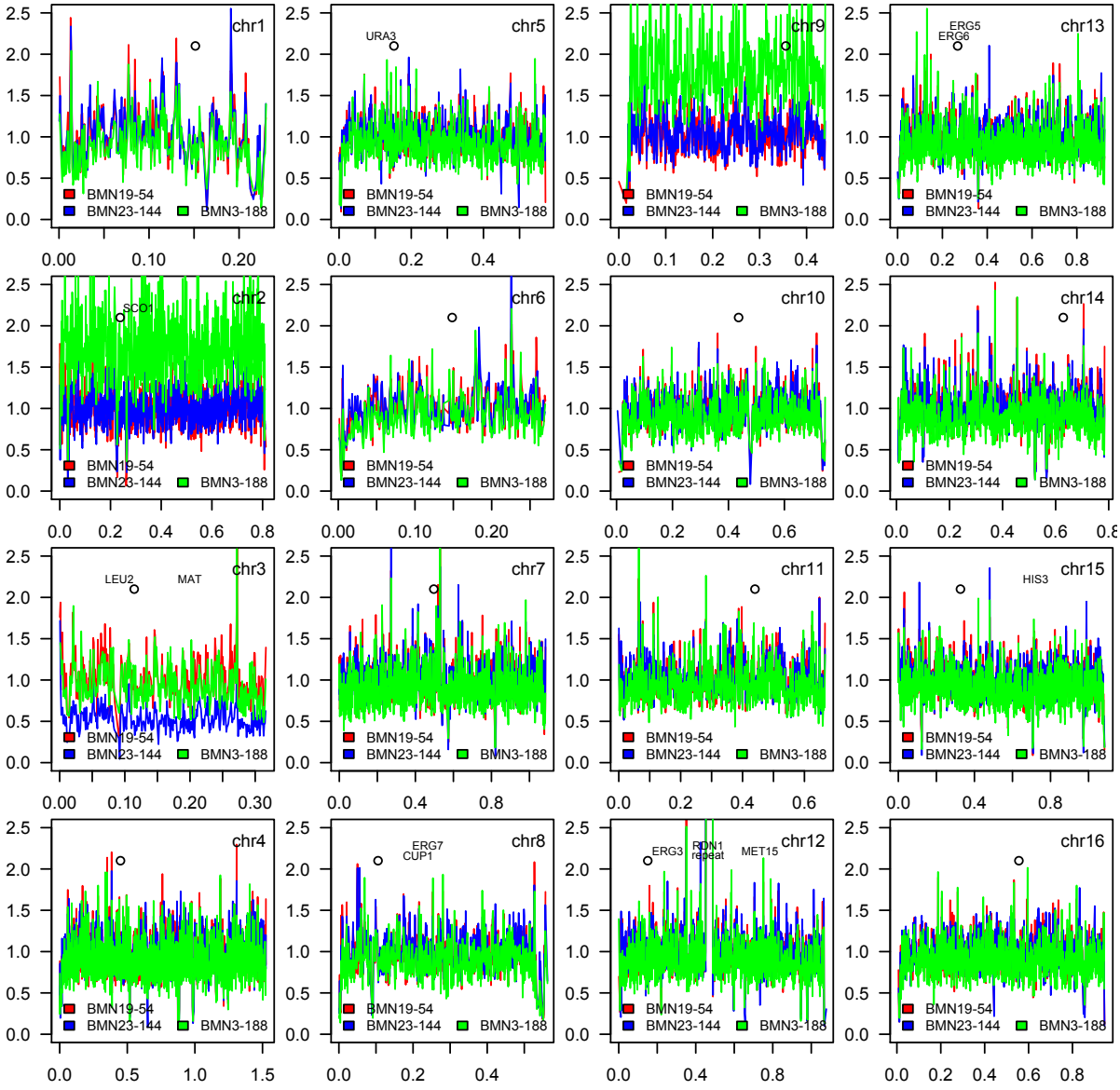


Figure E.4: Depth of coverage across the genome. For each chromosome in each line, alignments were obtained using Illuminas CASAVA-1.8.0 package. The pileup program in samtools-0.1.7a was then used to obtain reads by position across the genome. A custom perl script was then used to average depth of coverage in 1000 basepair windows across each chromosome (chr1 chr16). These windows skipped over sites where the depth of coverage was zero, which could be due to either deletions or ambiguous alignments (plots showing positions with no alignments are available upon request). Panel A: Lines BMN9-12, BMN9-130 and BMN28-14. Panel B: Lines BMN3-188 (note extra copy of chromosomes 2 and 9), BMN19-54, and BMN23-144 (note missing copy of chromosome 3).

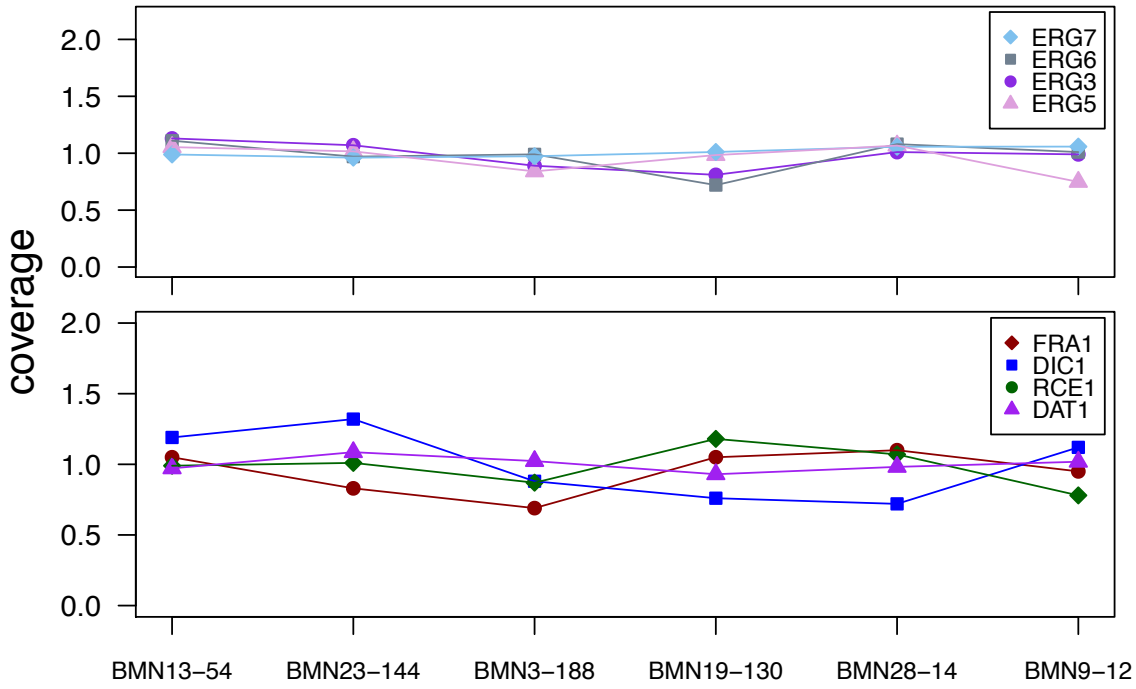


Figure E.5: Gene copy coverage for ergosterol and control genes from biPCR. Fragment counts from the beginning and end of each gene obtained by biPCR. The unix command "grep" was used to count the number of copies of the first and last 18bp of each ERG gene, in forward and reverse orientations, directly within the fastq sequences of unaligned 100bp paired end fragments obtained from Illumina sequencing. The average number of fragment counts was then divided by the average coverage across the genome from the CASAVA alignments. This relative fragment number is plotted, with "1" on the y-axis representing the median across the 6 lines (to minimize differences in scale across the genes). Panels show the relative fragment number from biPCR of the (top) ERG genes and (bottom) four control genes chosen from the same chromosomes as ERG3 and ERG6. For each gene, we confirmed by BLASTing the yeast genome (http://www.yeastgenome.org) that the 18bp at the beginning and end of each gene was unique.