Pathways of Adaptation:

A large-scale expression and genotypic analysis of the influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source.

By

Amber Lynne Scott

Bachelor of Science, The George Washington University, 2009

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Molecular, Cellular, and Developmental Biology

2017

This thesis entitled:
Pathways of Adaptation: A large-scale expression and genotypic analysis of the influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source.

Written by Amber Lynne Scott.
Has been approved by the
Department of Molecular, Cellular, and Developmental Biology

_____

Tom Blumenthal, Ph.D. (Chair)

_____

Robin Dowell, Sc.D. (Advisor)

Date_____

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

ABSTRACT

Scott, Amber Lynne (Ph.D., Molecular, Cellular, and Developmental Biology)

Pathways of Adaptation: A large-scale expression and genotypic analysis of the influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source.

Dissertation directed by Professor Robin Dowell.

Polyploidy, or having more than 2 full sets of chromosomes, has occurred in the evolution of many fungi, plant, and animal species and is thought to contribute to speciation. Despite the important role of polyploidy in evolution, little is known about how polyploidy contributes to adaptation and speciation. We previously showed that tetraploid yeast adapted significantly faster to growth under carbon stress compared to the haploid and diploid yeast, yet it was not clear what mechanisms drove the increase in the rate of evolution in the tetraploid strains. To answer this, I assessed the different pathways of adaption that the haploid, diploid, and tetraploid yeast strains took to adapt to growth in raffinose medium. I examined the molecular mechanisms of adaptation utilized by the strains through whole genome sequencing and RNA expression analysis of over 100 evolved clones. The evolved clones gained adaptive mutations in a narrow set of genes involved in glucose sensing and uptake, however the higher ploidy strains gained significantly different types of mutations than haploid strains. Additionally, I demonstrate that a gene expression signature of just 5 genes can accurately predict the gene that carries an adaptive mutation in the evolved clone. While many of the adaptive mutations occur in genes that encode proteins with known roles in glucose sensing and uptake, I also discover mutations in genes with no canonical role in carbon utilization

(*IPT1* and *MOT3*) as well as identify novel dominant mutations in glucose responsive regulators thought to only accumulate recessive mutations in carbon limited environments (*MTH1* and *RGT1*). I conclude that polyploid cells explore more genotypic and phenotypic space than lower ploidy cells. This is evidenced by a greater spectrum of beneficial mutations, including novel dominant alleles, copy number variation, and whole chromosome aneuploidy.

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Robin Dowell for her support and guidance through my graduate career. She is a dedicated mentor, a terrific role model, and a fierce advocate for women. I would like to thank all members of the Dowell lab past and present for helping me realize my talents as a molecular biologist as well as computational biologist. When I entered graduate school, I was unaware of the world of computational biology or that my skills and education were well suited to it. In particular, I would like to thank David Knox for teaching me python, Phillip Richmond for introducing me to bioinformatics tools, and Joey Azofeifa for continuing my computational education. I also want to thank fellow MCDB graduate students Tim Read and Jess Vera for all the help they gave me over the years. Mary Allen has been a great mentor and an inspiring scientist to work with over the years; her enthusiasm for science is contagious. This project would not have been possible without my collaborator, Anna Selmecki. She performed the initial experiments on which my thesis is based. Beyond that, she has been a friend, mentor, and enthusiastic cheerleader.

I would like to thank my thesis committee: Dr. Tom Blumenthal, Dr. Greg Odorizzi, Dr. Ken Krauter, and Dr. Mark Johnston. They provided great advice and supported me during my graduate career. In particular, I would like to thank Greg Odorizzi for introducing me to the wonder world of yeast during my rotation in his lab; he has continued to provide me with plasmids, strains, and advice in yeast biology throughout graduate school. I would also like to the Mark Johnston for his helpful advice about the nature of the mutations we recovered in our study as well as providing the tools to sequence a larger number of evolved clones and characterize the particular

mutants. Ken Krauter is a fierce advocate for graduate students and my graduate career would not have been the same without his guidance. Lastly, I would like to thank Tom Blumenthal for asking me the tough questions and always motivating me to work harder, better, and smarter for science.

I would like to thank the MCDB and BioFrontiers faculty and staff for their help over the years. Additionally, I would like to acknowledge the CU-Boulder and CU-Denver High-Throughput Sequencing Cores as well as the BioFrontiers Computing Core at the University of Colorado, Boulder for providing high performance computing resources (NIH 1S10OD012300) supported by BioFrontiers IT. This dissertation would not have been possible without funding by the Creative, Training in Molecular Biology NIH training grant (T32 GM007135), Blumenthal Fellowship, and Linda Crnic Institute graduate student fellowship. This research was supported by Creighton University, LB692-Nebraska Tobacco Settlement Biomedical Research Development New Initiative Grant, and NSF Career award (NSF 1350915).

Finally, I am extremely grateful to my friends and family without which the trials of graduate school would not have been possible. My fellow graduate students have provided me the constant support, steadfast friendship and a much-needed laugh at times. In particular, I would like to acknowledge Christa Trexler and Brittany Demmitt who have been my scientific sisters and my cheerleaders throughout graduate school. I would like to thank my family for years of encouragement and love. I especially want to thank my parents, Toni and Doug Sorenson, for motivating me to constantly achieve my goals and showing me that I am worthy of all my dreams. Lastly, I want to thank my

husband, David Scott, for his unwavering love and reassurance, this dissertation would not have been possible without his steady hand to support me.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1 INTRODUCTION

1.1 Scope of Thesis Research

The primary goal of this research is understand the identity and nature of adaptive mutations acquired by yeast of differing ploidy levels in evolution. This work is part of a larger collaboration with Anna Selmecki in David Pellman's lab, in which we performed a large-scale experimental evolution study with isogenic haploid, diploid, and tetraploid yeast grown in glucose-limiting conditions. My thesis has focused on utilizing evolved strains arising from the evolution study to investigate the molecular mechanisms of adaptation in the haploids, diploids, and tetraploids. In this introduction I will describe the history of evolution experiments, the influence of polyploidy in evolution and disease, and the functional changes that arise from whole genome duplication (WGD). Furthermore, I summarize experimental evolution studies that have examined the impact of ploidy level in adaptation. Since we evolved the yeast in glucose-limiting conditions, I will also briefly describe glucose sensing and uptake in yeast. Finally, I will summarize the other results obtained from the experimental evolution study performed by Anna Selmecki, namely the impact of polyploidy on the rates of adaptation.

1.2 The History of Experimental Evolution

Since the days of Darwin, evolutionary biologists have primarily studied descent with modification by comparing closely related living species, such as Darwin's finches, and inspection of the fossil record (Darwin 1859; Ridley 1983). A large wealth of knowledge was gained by these comparative studies throughout the greater part of the 19^{th} and 20^{th} centuries, especially when combined with phylogeny and statistical

methods (Harvey and Purvis 1991). However, comparative studies rely on correlation and are therefore generally unable to test hypotheses directly (Garland and Adolph 1994). As a result, scientists sought methods that would allow them to follow evolution over several generations, with controls and replication.

Experimental evolution studies utilize rapidly proliferating organisms in controlled environments to study the mechanisms of evolution. The first recorded experimental evolution study was performed by Reverend W. H. Dallinger in 1880, in which he evolved bacteria over the course of 7 years to withstand heat up to 158°F (Hass 2000). The utility of microorganisms in the study of experimental evolution was again realized in the 1950s, when scientists observed that *Escherichia coli* under constant or periodic selection continually produced a fitter phenotype (Atwood et al. 1951; Ryan 1953). At the same time, *Drosophila* were used to study genetic drift and natural selection in the lab (Dobzhansky and Pavlovsky 1957; Futuyama 1970). *Drosophila* would remain the primary tool in experimental evolution studies throughout the latter half of the 20th century (Elena and Lenski 2003).

1.3 Experimental Microbial Evolution

Recent decades have seen a resurgence in experimental evolutionary studies using microorganisms such as bacteria, yeast, and even viruses. These organisms are ideal for experimental evolution studies because they have a short generation time, allowing researchers to study hundreds of generations of evolution in real time. In the 1950s Francis J. Ryan noted that “in two years bacteria can grow through more generations than man has in the millions years of so of his evolutionary history” (Ryan 1953). Microorganisms are small and easy to grow and maintain, allowing for

manipulation of environmental variables, such as salinity (Bell and Gonzalez 2009), nutrient availability, such as carbon and nitrogen (Boer et al. 2003), or chemical stressors, such as fungicide (Cowen et al. 2000). Importantly, microorganisms can be stored in suspended animation indefinitely allowing for the direct comparison of ancestors to evolved populations. Furthermore, microorganisms are ideal for testing the reproducibility of evolutionary outcomes from the same ancestor (Ferea et al. 1999; Cooper et al. 2003) as well as the effect of the organism's genetic background on adaptability (Travisano et al. 1995).

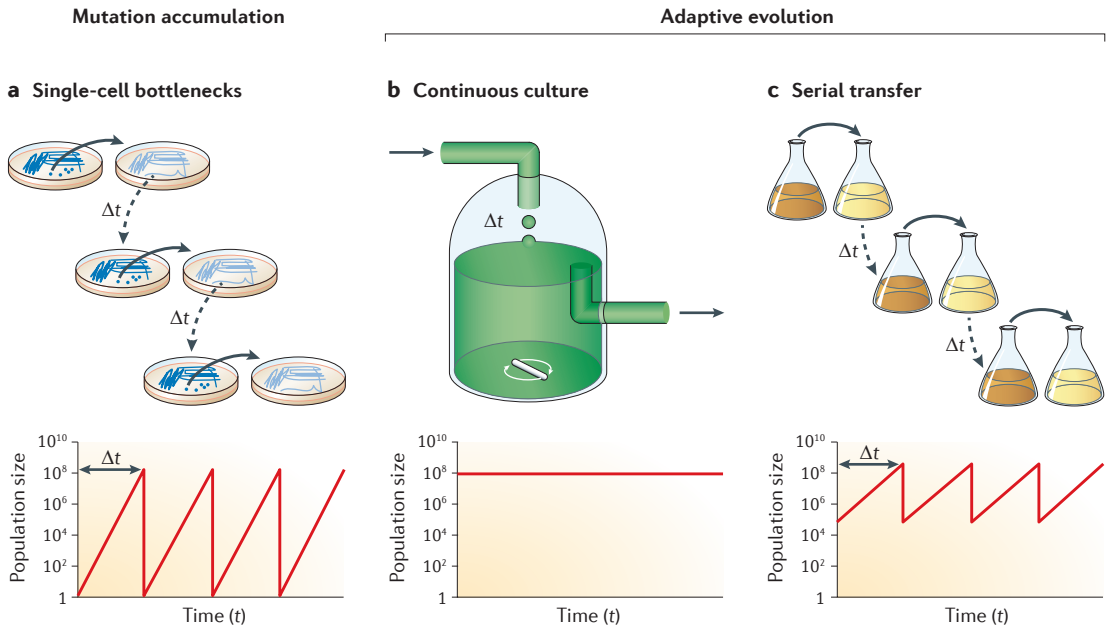


Figure 1-1 Methods of experimental evolution A graphical summary of the primary methods utilized in experimental evolution studies. A) Single-cell bottleneck experiments study the accumulation of mutations through repeated growth from a single colony on agar plates. B) Continuous culture experiments grow microorganisms in a constant environment and steady-state population size. Continuous culture experiments allow for adaptive evolution to a selective pressure, such as nutrient concentration. C) In serial transfer, also known as mass transfer, a portion of the population is transferred at regular intervals to fresh media. Similar to (B), serial transfer experiments study adaptive evolution under selective pressure. The figure is used with permission from Barrick and Lenski (2013).

There are three main techniques utilized to evolve microbes in the lab: single-cell bottlenecks, continuous culture, and serial transfer (Figure 1-1). Single-cell bottlenecks examine the accumulation of arbitrary mutations through repeated population restriction, by sequential growth of single colonies of microbes on agar plates, leading to the fixation of mutations with no fitness effect (Figure 1-1A). The single-cell bottleneck approach is useful for estimation of mutation rates, however successive bottlenecks lead to an overall decay in fitness over time (Barrick and Lenski 2013). Continuous culture and serial transfer experiments allow researchers to apply a selective pressure to the cultures in order to understand the effect of natural selection on populations. Continuous culture in a chemostat keeps the nutrient availability and population size constant over time through the in-flow of nutrients and out-flow of waste and excess individuals in the population (Figure 1-1B) (Barrick and Lenski 2013). Alternatively, in serial transfer experiments, a portion of the population is transferred to fresh media, favoring mutations that increase exponential growth following transfer while still maintaining genetic diversity (Figure 1-1C) (Barrick and Lenski 2013). In our study, we evolved isogenic haploid, diploid, and tetraploid yeast in glucose-limiting conditions by serial transfer.

The genomics era has had a major influence on the renaissance of experimental microbial evolution. Whole genome sequencing of evolved clones has improved measurements of mutation rates and allowed researchers to determine the identity of adaptive mutations. Furthermore, deep sequencing of evolved populations, before and after evolution, has led to a greater understanding of population dynamics and genome evolution (Reviewed in Barrick and Lenski 2013). Recently, Levy et. al. tagged

~500,000 isogenic yeast with unique barcodes and tracked their individual lineages for ~168 generations in carbon-limited conditions (Levy et al. 2015). This was the first study in which researchers observed population dynamics on a single-cell level. As sequencing technology improves and costs decrease, it is feasible that single-cell whole genome sequencing in evolving populations will further improve the resolution of evolutionary dynamics and the discovery of beneficial mutations.

RNA sequencing and microarray have been used the measure global changes in genes expression to understand how strains and populations adapt to growth in different environments or under nutrient limitation. Expression analysis in evolved clones has shown evidence of major metabolic shifts that may be adaptive in nutrient-limited conditions (Ferea et al. 1999; Gresham et al. 2008). Moreover, evolved clones from independent populations exhibited similar gene expression patterns, evidence of parallel evolution (Ferea et al. 1999; Cooper et al. 2003; Gresham et al. 2008). In my thesis work, I integrate both whole genome sequencing and expression analysis to understand the molecular mechanisms of adaptation in the haploid, diploid, and tetraploid clones after growth in glucose-limited conditions.

1.4 The Significance of Polyploidy in Evolution

Polyploidy, or having greater than two full complements of chromosomes, has occurred in plant, animal and fungi lineages throughout the course of evolution and continues to persist in some of these lineages today (Otto and Whitton 2000). With the advent of whole genome sequencing, has become possible to reconstruct the past on a genomic level to look for evidence of whole genome duplication events (WGD) that occurred during a species' evolution (Van de Peer 2004). There is evidence to support

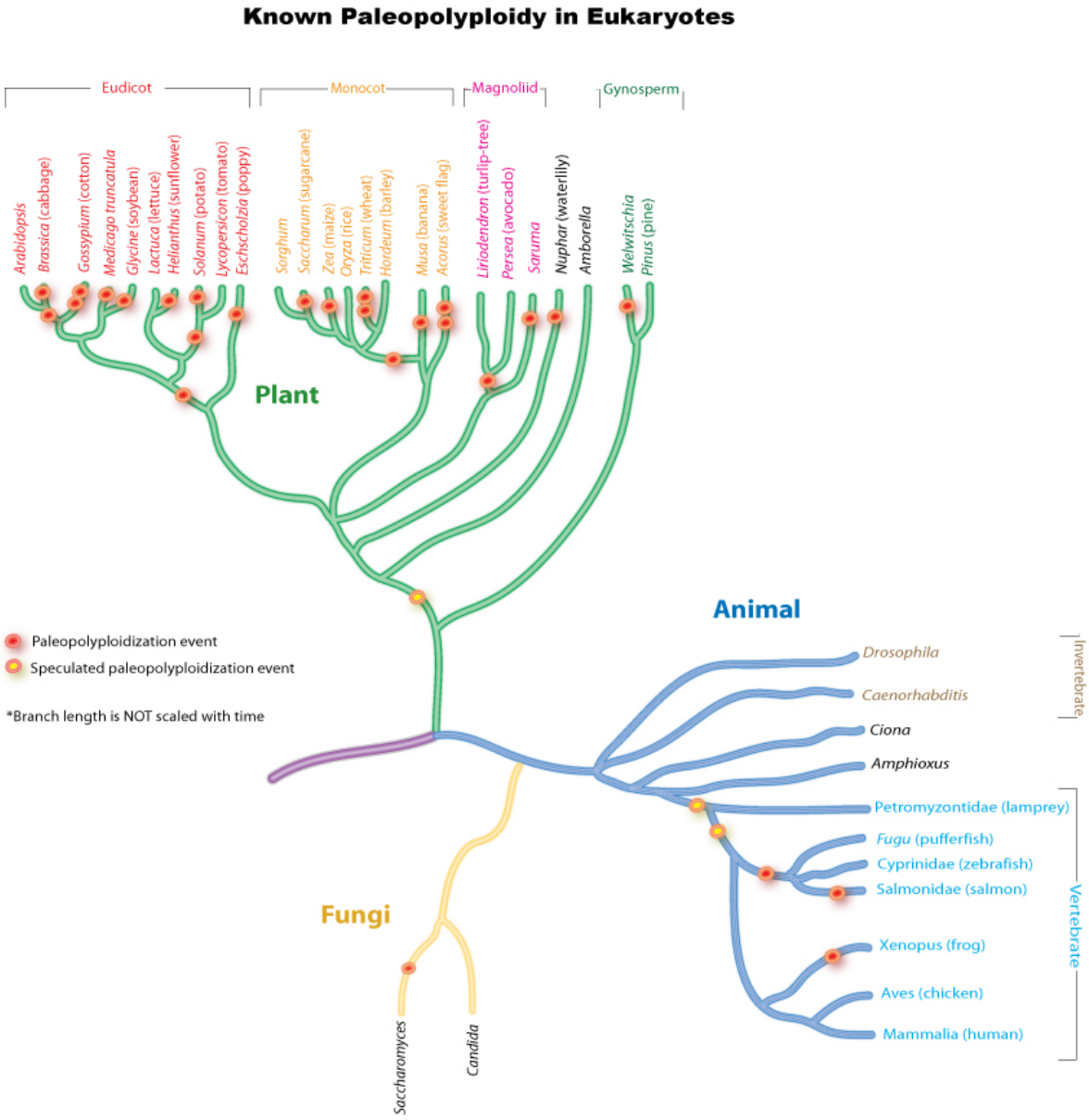


Figure 1-2 Paleopolyploidy in eukaryotes Paleopolyploidyzation events occurred in the evolution of plants (green), animals (blue), and fungi (yellow). Well-supported whole genome duplication events are indicated by a red dot. Yellow dots indicate speculated whole genome duplication events. Branch lengths are not to scale. Data compiled from (Wolfe 2001; Adams and Wendel 2005; Cui et al. 2006) by Peter Zhang (Wikipedia) and the figure is reproduced here under the creative commons license.

ancient polyploidization events in plants, animals and fungi (Figure 1-2)(Otto and Whitton 2000; Wolfe 2001; Adams and Wendel 2005; Cui et al. 2006). If polyploidy were an evolutionary dead-end, we would expect few species would result after the WGD event, however this is not what is observed (Figure 1-2). Instead, polyploidization events occur early in lineages, thus they are likely to contribute to evolutionary success (Otto 2007).

The frequency of polyploidy arising in plants and animals is surprisingly high. Plants, the most common polyploids, produce non-reduced gametes at a rate of 0.5% per gamete and produce polyploid offspring at rates of 6.3% (Ramsey and Schemske 1998). Additionally, 0.9% of chicken embryos are triploid or tetraploid and 5.3% of spontaneous human abortions are triploid or tetraploid (Bloom 1972; Creasy et al. 1976). However, polyploidy is almost always fatal in birds and mammals (Otto and Whitton 2000). Polyploidy is also very common in fungi. Whole genome sequencing of 132 clinical isolates of *Saccharomyces cerevisiae* identified that ~34% are triploid or tetraploid. A study of wild yeast in “Evolution Canyon” found that 69% of strains isolated were triploid or tetraploid, compared to 31% that were diploid (Ezov et al. 2006; Zhu et al. 2016). The high frequency of polyploidy suggests it is an important source of genomic variation and is likely to have a role in evolution.

A related major driving force in evolution is gene duplication, which is considered a primary determinant of increased organismal complexity (Ohno 1970; He and Zhang 2005). However, gene duplication is relatively rare, while polyploidy and aneuploidy are much more common (Lynch and Conery 2000; Otto 2007). Thus, WGD events produce the raw material (i.e. large scale duplications) on which evolution can act to produce a

new gene function. Relaxed selection on duplicated genes allows for the evolution of sub-functionalization, such as the hexose transporters in yeast that have differing affinities for glucose, or neofunctionalization, such as the evolution of glucose sensors from glucose transporters in yeast (Force et al. 1999; Lynch and Force 2000; Sémon and Wolfe 2007a; Lin and Li 2011). In fact, WGD events in the evolution of *Arabidopsis thaliana* are responsible for >90% of the increase in transcription factors, signal transducers, and developmental genes in their genome (Maere et al. 2005). Despite the high probability of inactivation of duplicated genes, duplicated genes following WGD are preserved at a rate of 30 to 50% (Lynch and Conery 2000).

Polyploidy may also contribute to evolution by increasing the rate of adaptation by doubling the target size for beneficial mutations (Adams and Hansche 1974; Otto 2007; Gerstein and Otto 2009). In smaller populations the rate of generating mutations is limiting, therefore the increased target size of polyploid cells relative to haploids is predicted to increase the rate of adaptation (Orr and Otto 1994; Otto and Whitton 2000; Zeyl et al. 2003; Otto 2007). However, beneficial phenotypic changes may be 'masked' in higher ploidy cells depending on the degree of dominance of the mutations (Orr and Otto 1994; Otto and Whitton 2000; Anderson et al. 2004). Increased numbers of chromosome sets may also buffer the effects of deleterious mutations, but this is only a temporary advantage, as it leads to the accumulation of deleterious mutations in the genome with diminishing returns (Korona 1999; Otto and Whitton 2000; Thompson et al. 2006). Thus, in certain environments WGD may provide temporary benefit long enough for evolution to act.

1.5 Polyploidy and Disease

Polyploidy not only contributes to evolution, but is also an important factor in human disease. Polyploid cells are formed in the human body in a variety of pathological conditions (Storchova and Pellman 2004). For example, polyploid cells arise as a result of aging, hypertension, and wound healing (Oberringer et al. 1999; Hixon et al. 2000; Gorla et al. 2001). Tetraploidy is considered an intermediate on the road to tumorigenesis, particularly in the development of aneuploidy, an unequal number of chromosomes (Ganem et al. 2007). 37% of all cancers undergo a genome duplication event sometime during their progression (Zack et al. 2013). Additionally, in a mouse model of breast cancer, polyploid cells promote tumorigenesis, whereas diploid cells do not (Fujiwara et al. 2005). The prevalence of polyploidy in cancer and other human disease suggests that polyploidy may be beneficial under certain conditions, such as metabolic stress (Storchova and Pellman 2004). Polyploidy is also frequently identified in human fungal pathogens, including those that cause life-threatening infections like *Cryptococcus neoformans*, *Candida albicans,* and *Saccharomyces cerevisiae* (Clemons et al. 1997; Muller and McCusker 2009; Selmecki et al. 2010; Harrison et al. 2014; Zhu et al. 2016). Despite the importance of polyploidy in evolution and human disease, little is known about how increasing ploidy levels affect adaptation to a stressful environment on a molecular level.

1.6 The Functional Effects of Polyploidy

Polyploid cells exhibit unique features compared to haploid and diploid cells such as an increased cell size and, as a result, a decreased surface area to volume ratio (Galitski et al. 1999; Mable 2001; Cook and Tyers 2007; Lee et al. 2010). Furthermore,

in some environments these differences in cells size can affect growth rate and nutrient transport (Adams and Hansche 1974; Weiss et al. 1975; Mable 2001). Gene expression alterations have been reported in some polyploids, particularly polyploids formed by the hybridization of different species (known as allopolyploids) (Adams and Wendel 2005; Chen and Ni 2006). However, since increasing ploidy levels maintains gene dosage balance, polyploids formed by the same species (autopolyploids) exhibit few gene expression changes (Birchler et al. 2001; Storchová et al. 2006; Wang et al. 2006). To fully understand the impact of ploidy level on gene expression, the Fink lab constructed isogenic haploid, diploid, triploid and tetraploid yeast strains, as well as isogenic cell size mutants. They conclude that the primary gene expression differences with increased ploidy are the result of cell size alterations, consistent with a deceased surface area to volume ratio (Galitski et al. 1999; Wu et al. 2010).

Another major feature of polyploid cells is problems with mitosis and meiosis, and cells often undergo rapid genomic rearrangement following WGD (Wendel 2000; Adams and Wendel 2005; Comai 2005; Otto 2007). In particular, there is evidence of genomic instability following ancient genome duplication events (Sémon and Wolfe 2007b). Newly formed polyploids exhibit increased transposable element activity that results in chromosome restructuring (Wendel 2000). Polyploidy also leads to chromosomal instability, namely aneuploidy (Mayer and Aguilera 1990; Storchova and Pellman 2004). This chromosome instability may be beneficial in some conditions, as aneuploidy itself has been shown to be beneficial in certain environments (Selmecki et al. 2006; Rancati et al. 2008; Pavelka, Rancati, and Li 2010). On the other hand, aneuploidy is most often thought to be detrimental to fitness (Pavelka, Rancati, Zhu, et al. 2010; Oromendia et al.

2012; Sunshine et al. 2015) and polyploids often return to near diploid levels over time through chromosome loss (Gerstein et al. 2006). However, tetraploid yeast that were adapted in non-stress conditions (rich media at room temperature) for 1000 generations remained 4N and exhibited greater genome stability under stress than newly formed tetraploids, suggesting that polyploidy is only unstable in certain environments and once established can remain stable (Lu et al. 2016).

1.7 Studies of Ploidy Level in Experimental Evolution

Ploidy level changes, and in particular polyploidy, represent one of the most important means in which an organism can generate large-scale genotypic and phenotypic variation (King et al. 2012; Soltis et al. 2014). However, witnessing spontaneous polyploidization events and following their evolutionary trajectories is difficult in nature, but it is possible in laboratory-controlled experiments with single-celled organisms. Many *in vitro* evolution studies have been performed with *S. cerevisiae* and *Candida albicans*, yet most of these focus on haploid and diploid cells (summarized Table 1-1).

Theoretical models suggest that increased ploidy level increases the rate of adaptation (Stebbins 1940; Otto and Whitton 2000). However, the literature has conflicting evidence about the relative rate of adaption or the fixation of beneficial mutations in haploid and diploid *S. cerevisiae.* An early study that examined the rate of adaptation through the fixation of beneficial mutations in haploid and diploids grown in carbon-limited chemostats found that diploid populations fixed beneficial mutation at 1.6x the rate of haploid populations (Paquin and Adams 1983). Yet, another study that examined the rate of adaptation of isogenic haploid and diploid yeast found that haploid

Ploidy Levels	EE Conditions	Ploidy Level Conclusions	Citations
1N, 2N	Chemostat, low-glucose, 300 generations, *S. cerevisiae*	Diploids adapt/fix mutations 1.6x faster than haploids	(Paquin and Adams 1983)
1N, 2N	Serial transfer, glucose-limiting, 30C, *S. cerevisiae,* 5000 generations, large & small population size	Haploids adapt faster in large populations, and diploid and haploids adapt at the same rate in small populations	(Zeyl et al. 2003)
1N, 2N	Serial transfer, Fluconazole, 30C, 400 generations, *C. albicans*	Haploids adapted faster at high drug concentrations, but diploids fixed mutations faster at lower concentrations, depends on dominance	(Anderson et al. 2003; Anderson et al. 2004)
1N, 2N	Serial transfer, Rich media, non-fermentable, low-glucose, high salt, 350 generations, mutator vs non- mutator, *S. cerevisiae*	Diploids mask deleterious mutations and this increases competitive fitness	(Thompson et al. 2006)
1N, 2N, 4N	Serial transfer & mutation accumulation, YPD & NaCl, 30C, 1800 generations, *S. cerevisiae*	1N and 4N populations converge on diploidy, dependent on environment and initial ploidy, diploidization rarely occurred in mutation accumulation	(Gerstein et al. 2006; Gerstein et al. 2008)
1N, 2N	Chemostat, 3 environments (low-glucose, sulphate, phosphate), S288c and CEN.PK, ~200 generations, *S. cerevisiae*	Diploids more likely to gain chromosome-scale CNVs	(Gresham et al. 2008)
1N, 2N	Chemostat, low-glucose, 450 generations, *S. cerevisiae*	CNV is first mutation in 2N, CNVs are over-dominant	(Kao and Sherlock 2008; Wenger et al. 2011; Kvitek and Sherlock 2013; Sellis et al. 2016)
1N, 2N	Serial transfer, 6 environments (Caffeine, EtOH, HCL, KOH, NaCl, Naystatin), 187 generations, *S. cerevisiae*	Haploids adapt faster in a range of environments, affected by level of dominance	(Gerstein et al. 2011)
1N, 2N, 4N	Turbidostat, EtOH, 200 generations, *S. cerevisiae*	Convergence on diploid state in all populations, high levels of aneuploidy, mutator phenotype, parallelism in mutations	(Voordeckers et al. 2015)
2N, 4N	Serial transfer, YPD, 23C, 1000 generations, *S. cerevisiae*	At 23C tetraploid evolved clones remain 4N, evolved greater stability at 30C, become "diploid like" in expression and phenotype	(Lu et al. 2016)
1N, 2N, 4N, 3N+	Serial transfer, 4 environments (Rich Media, Minimal Media, phosphorus-limited, nitrogen-limited), 30C, 140 generations, *Candida albicans*	Populations almost always converged on diploidy, confirming "ploidy drive"	(Gerstein et al. 2017 Feb)

Table 1-1 Summary of experimental evolutions studies on ploidy level

populations adapted faster in the majority of conditions tested (Gerstein et al. 2011). These conflicting studies suggest that the extent of the difference in the rate of adaptation between haploids and diploids depends on the specific condition and the level of dominance of the mutations. In support of this, studies that experimentally evolved *C. albicans* in differing concentrations of the antifungal drug Fluconazole demonstrated that haploid cells adapted faster in an environment where recessive mutations were favored, and diploid cells did better in an environment that required dominant mutations. (Anderson et al. 2003; Anderson et al. 2004). Population size has also been shown to affect the rate of adaption, haploid cells evolved faster than diploid cells in large populations, but when the population size was reduced there was no advantage to haploidy (Zeyl et al. 2003).

The only experimental evolution studies to directly compare haploid, diploid, and tetraploid yeast focused on “ploidy drive”, or the pressure to return to the natural ploidy level. Haploid and tetraploid *S. cerevisiae* strains grown in YPD, high salt, and ethanol converge on 2N, while diploids remain primarily 2N (Gerstein et al. 2006; Gerstein et al. 2008; Voordeckers et al. 2015). Similarly, laboratory and clinical isolates of *C. albicans* also demonstrated a strong “ploidy drive” (Gerstein et al. 2017). In all cases, the rate of convergence was dependent on the environment and initial ploidy level. However, the above studies were all performed at 30°C, which has been shown to cause stress in tetraploid yeast (Storchová et al. 2006). In a recent study, tetraploid yeast grown in rich media for 1000 generations at 23°C remained largely 4N (Lu et al. 2016). The evolved 4N clones showed a gene expression signature similar to ancestral diploids and have

greater genome stability than the ancestral tetraploids when grown at 30°C, suggesting that tetraploid cells can evolve to maintain genome stability in the right conditions.

To better understand the molecular mechanisms of evolution, particularly after whole genome duplication, more studies are needed that directly compare the rate of adaptation as well as the genetic and phenotype changes in haploid, diploid, and polyploid cells during evolution. Dr. Anna Selmecki, we performed *in vitro* evolution of isogenic haploid, diploid, and tetraploid yeast in raffinose media, a carbon-limited environment. We further monitored the rate of adaptation in hundreds of parallel populations and quantified the genomic and phenotypic changes following adaptation.

1.8 Glucose Uptake and Metabolism in Yeast

For these studies we utilized raffinose as the sole carbon source in the growth medium. Raffinose is hydrolyzed extracellularly into fructose and melibiose; the fructose is then metabolized in yeast by the glucose pathway (Lagunas 1993). Melibiose cannot be further metabolized in our lab strain because it lacks key melibiase enzymes (Naumov et al. 1990). Thus, raffinose medium is a surrogate for low-glucose growth medium. However, unlike low-glucose media in which the glucose is quickly depleted from the environment, raffinose is metabolized extracellularly, thereby serving as a constant source of low carbon in the environment (Ozcan et al. 1996).

A great deal is known about the glucose pathway and extensively reviewed in (Johnston 1999; Ozcan and Johnston 1999; Rolland et al. 2002; Gancedo 2008); here we provide a brief overview. There are three key steps that lead to glucose/fructose uptake in yeast: the glucose sensors, signal transducers, and hexose transporters (Figure 1-3). The glucose sensors, Snf3p and Rgt2p, signal to downstream regulators of

glucose responsive genes in the presence of low or high extracellular glucose concentrations, respectively (Ozcan et al. 1996; Sabina and Johnston 2009). Mth1p binds Rgt1p and promotes Rgt1p mediated repression of the glucose responsive genes (Flick et al. 2003; Moriya and Johnston 2004). Low levels of extracellular glucose activate a signaling cascade, through Snf3p, that results in the degradation of Mth1p (Ozcan et al. 1996; Flick et al. 2003). Loss of Mth1p relieves the Rgt1p repression of glucose responsive genes, such as the hexose transporters and the invertase, *SUC2,* which extracellularly hydrolyzes raffinose into melibiose and fructose (Figure 1-3) (Carlson and Botstein 1982; Ozcan et al. 1996; Polish et al. 2005).

There are 17 annotated hexose transporters in yeast, however, only Hxt1p-Hxt7p have been shown to transport glucose into the cell (Figure 1-3) (H Liang and Gaber 1996; Boles and Hollenberg 1997; Lin and Li 2011). The transporters differ in their affinity for glucose and their regulation. Hxt1p has the lowest affinity for glucose and is upregulated when there are high levels of glucose outside the cell; on the other hand Hxt2p and Hxt4p have high and moderate affinities for glucose, respectively, and are induced by low concentrations of glucose in the media (Kruckeberg and Bisson 1990; Theodoris et al. 1994; Boles and Hollenberg 1997; Reifenberger et al. 1997). Hxt3p has intermediate affinity for glucose and is upregulated by both low and high levels of extracellular glucose (Boles and Hollenberg 1997; Reifenberger et al. 1997). Hxt6p and Hxt7p have the highest affinity for glucose and high levels of basal expression (Boles and Hollenberg 1997). *HXT6* and *HXT7* are also 99.7% identical and therefore indistinguishable when measured by qRT-PCR. Snf3 is required for induction of *HXT2* and *HXT4,* but not the induction of *HXT6* and *HXT7* (Hong Liang and Gaber 1996).

Several experimental evolution studies in yeast have been performed in glucose limiting environments (Paquin and Adams 1983; Brown et al. 1998; Ferea et al. 1999; Gresham et al. 2008; Kao and Sherlock 2008; Koschwanez et al. 2013; Kvitek and Sherlock 2013; Levy et al. 2015). However, these studies were performed in haploid or diploid populations. Thus, evolution of haploid, diploid and tetraploid populations in raffinose media is a natural extension of previous experimental evolution studies in carbon-limited environments.

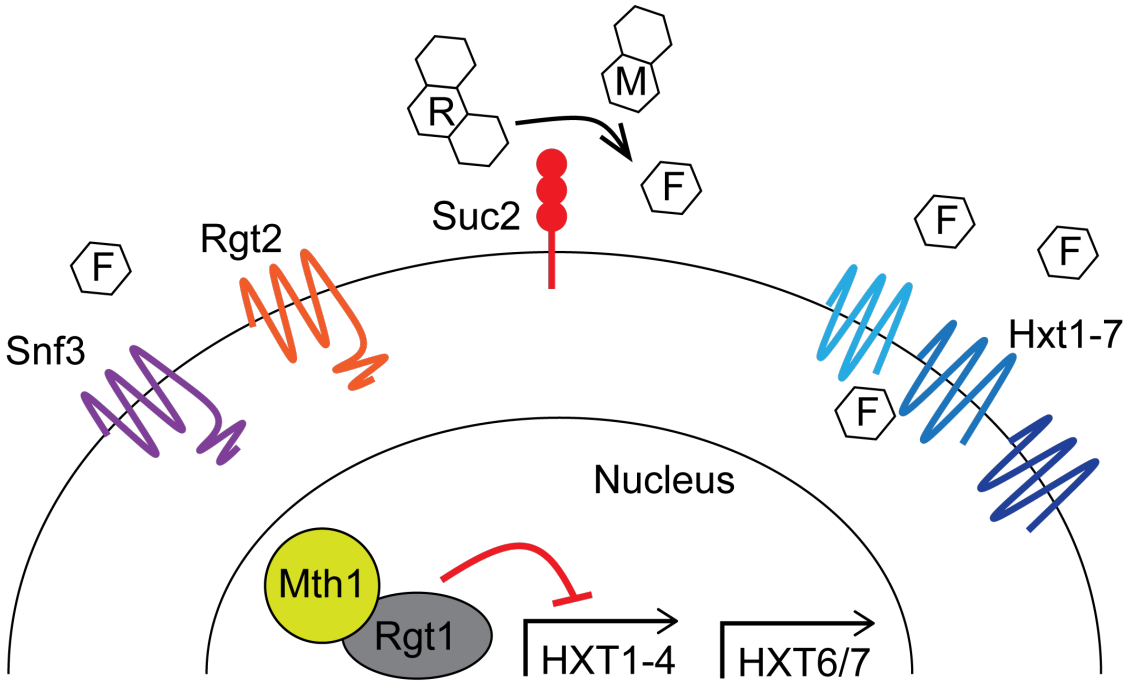


Figure 1-3 Raffinose metabolism in yeast Raffinose is extracellularly hydrolyzed into fructose and melibiose. Fructose can then be metabolized equivalently to glucose by the cell. Low levels of extracellular fructose activate signaling by the glucose sensor Snf3p, which leads to the degradation of Mth1p and phosphorylation of Rgt1p. Phosphorylation of Rgt1p blocks DNA binding of Rgt1p and results in the depression of glucose responsive genes.

1.9 Polyploidy Can Drive Rapid Adaptation in Yeast

The experimental evolution on which my thesis is based was performed by Anna Selmecki in a collaboration between David Pellman's and Robin Dowell's labs. My primary role in the initial publication was the sequencing of over 120 clones derived from independent haploid, diploid, and tetraploid evolved populations, work I describe in detail in Chapter 0. Below I describe the other key findings from this study, as motivation for my thesis. This initial study is published in:

Anna M. Selmecki, Yosef E. Maruvka, Phillip A. Richmond, Marie Guillet, Noam Shoresh, **Amber L. Sorenson**, Subhajyoti De, Roy Kishony, Franziska Michor, Robin Dowell & David Pellman. 2015. Polyploidy can drive rapid adaptation in yeast. Nature 519:349–352.

Polyploidy is proposed to alter the rate of evolutionary adaptation and may also affect adaptation independently of beneficial mutations through ploidy-specific changes in cell physiology. Anna Selmecki performed a large-scale *in vitro* evolution experiment to test directly whether polyploidy can accelerate evolutionary adaptation. Anna constructed isogenic haploid (1N), diploid (2N), and tetraploid (4N) yeast and performed experimental evolution by serial transfer for hundreds of independent populations (Figure 1-4A). Yeast were grown in synthetic complete media with raffinose as the sole source of carbon, a well studied stressor in yeast (Kruckeberg and Bisson 1990). To monitor the progress of adaptation, the evolution experiments were performed as competitions between equal numbers of cells of the same ploidy expressing cyan fluorescent protein (CFP) or yellow fluorescent protein (YFP) (Hegreness 2006; Kao and Sherlock 2008). Flow cytometry was performed daily to monitor the percentage of

YFP and CFP expressing cells in the population (Figure 1-4A). The acquisition and spread of beneficial mutations is visualized by divergence from a 50:50 ratio of CFP- and YFP-expressing cells (Figure 1-4B). The rate of adaptation was determined by measuring the change in fitness relative to the diploid ancestor over time.

The evolution experiment demonstrated that, over 250 generations, the tetraploids adapted at a rate that was significantly faster than haploids or diploids (Figure 1-4C; t-test, $P<1x10^{-10}$). Additionally, mathematical modeling indicated that in a glucose-limited environment polyploidy increases the rate (Figure 1-4C) and fitness effects of the acquired mutations (Figure 1-4D). To further understand the molecular basis of adaptation and the role of ploidy level in adaptation to rapid growth in raffinose media, I performed whole genome sequencing and expression analysis in over 100 evolved clones (described in Chapter 2). I found increased genetic complexity in polyploid evolved clones relative to haploids and diploids, suggesting that in this relatively short evolutionary timescale polyploid cells explored more evolutionary innovations. Additionally, I focused on the gene expression patterns acquired during *in vitro* evolution (described in Chapter 3) to determine if polyploid cells have increased variation in gene expression and regulatory wiring (Osborn et al. 2003). Interestingly, gene expression patterns for all evolved clones cluster according to one key adaptive mutation that they carry, despite additional background mutations and underlying karyotype or ploidy level. I further characterize the key adaptive mutations gained in adaption to rapid growth in raffinose media (described in Chapter 4). Overall, this work suggests that polyploid cells can rapidly adapt to a novel environment due to an increased sampling of key adaptive mutations.

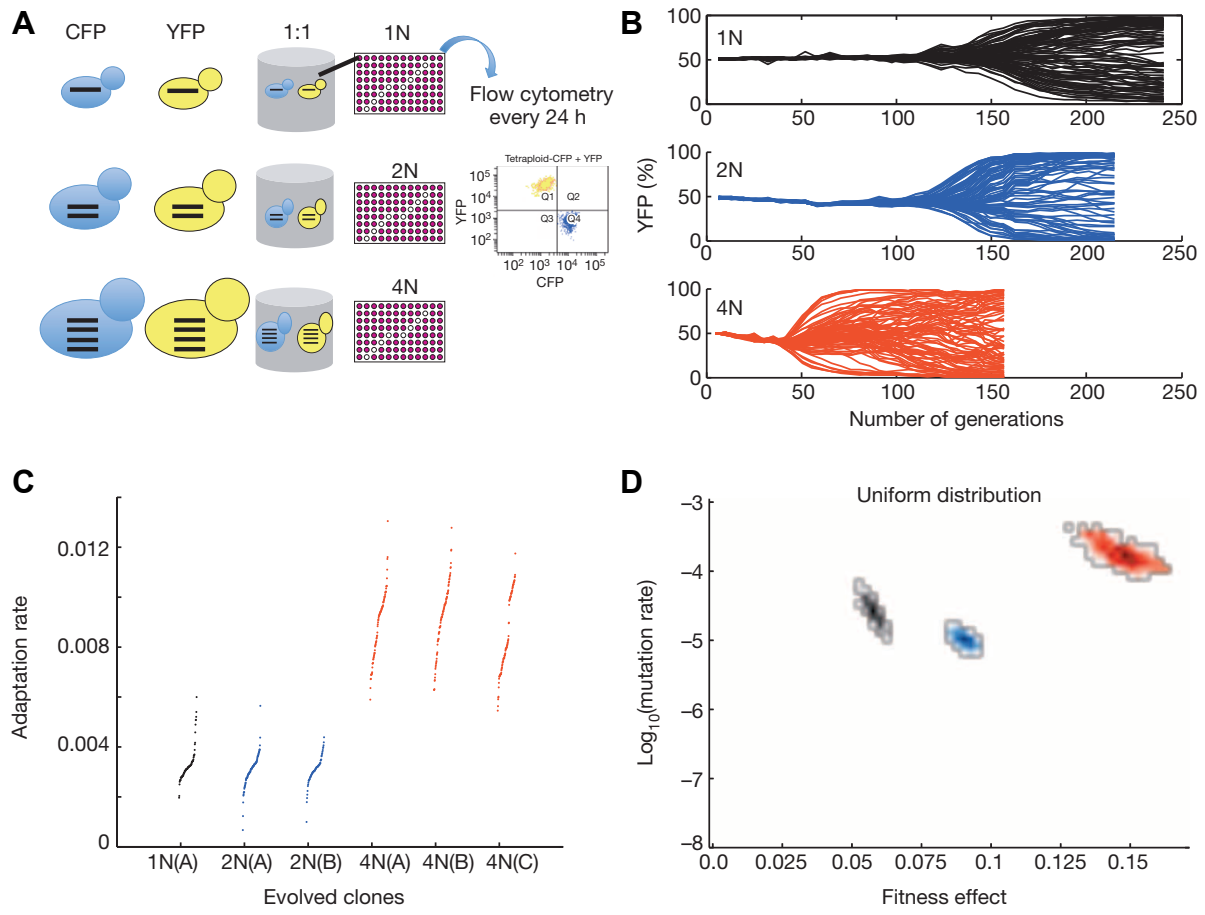


Figure 1-4 Rapid spread of beneficial mutations in tetraploid yeast. A) Schematic diagram of the evolution experiment. B) Flow cytometry analysis of isogenic haploid (black), diploid (blue), and tetraploid (red) populations during adaptation to raffinose medium. Each line is the percentage of YFP cells in an independent population of YFP and CFP cells. Data from haploids is black, from diploids is blue, and from tetraploids is red. C) The adaptation rate of the evolved clones relative to the diploid ancestor after 250 generations. Data points are the average rate of adaptation (change in fitness between generation 250 and generation zero, divided by 250 generations) of two replicate fitness measurements for the evolved clones. Clones from replicate evolution experiments (A, B, and C) are indicated. The tetraploids acquired significantly more fitness in the same number of generations compared with the haploids and diploids (t-test, $P < 1x10^{-10}$). D) Estimates from the branching evolution model of the best-fit value of the selection coefficient and beneficial mutation rate of each ploidy experiment, and their error range, determined using a uniform distribution of acquired mutations. Error ranges were obtained by parametric bootstrap of 1,000 independent realizations. For detailed methods on the evolution experiments and modeling see Section 8.3. Figure and legend from (Selmecki et al. 2015).

2 MUTATIONAL SPECTRUM VARIES BY PLOIDY LEVEL

Portions of this chapter are published previously or are currently under review:

1. Anna M. Selmecki, Yosef E. Maruvka, Phillip A. Richmond, Marie Guillet, Noam Shoresh, **Amber L. Sorenson**, Subhajyoti De, Roy Kishony, Franziska Michor, Robin Dowell & David Pellman. 2015. Polyploidy can drive rapid adaptation in yeast. Nature 519:349–352.
2. **Amber L Scott**, Phillip A. Richmond, Robin Dowell, Anna M. Selmecki. The influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source. *Under Review.*

Anna Selmecki performed the experimental evolution experiment and comparative genome hybridization assays (aCGH). Phillip Richmond and Robert Thomas performed sequence alignment and variant calling in the whole genome sequencing. For complete details see the materials and methods (Chapter 2.3).

2.1 Introduction

We passaged isogenic haploid (1Ne), diploid (2Ne), and tetraploid (4Ne) yeast strains in raffinose medium, described in detail in the additional methods (Chapter 8.3). Raffinose, a poor source of carbon, provides a low but constant source of carbon in the media, which causes reduced growth rate relative to glucose. After roughly 250 generations a single clone was isolated from each evolved population and both RNA and DNA were isolated for further analysis (Figure 2-1). By convention, clones isolated from 1Ne, 2Ne, and 4Ne populations are denoted with strain identifiers in the 100s, 200s, and 300s, respectively.

To understand the molecular mechanisms that lead to the adaptation of the tetraploid lineages to raffinose, and how these differed from haploid and diploid lineages, I first quantified the number and types of mutations gained in the evolved clones after 250 generations. I utilized whole genome sequencing (WGS) and comparative genome hybridization microarray (aCGH) to detect single nucleotide variants (SNVs) and copy number variations (CNVs) in each of 24 evolved haploid, 24 evolved diploid, and 28 evolved tetraploid clones.

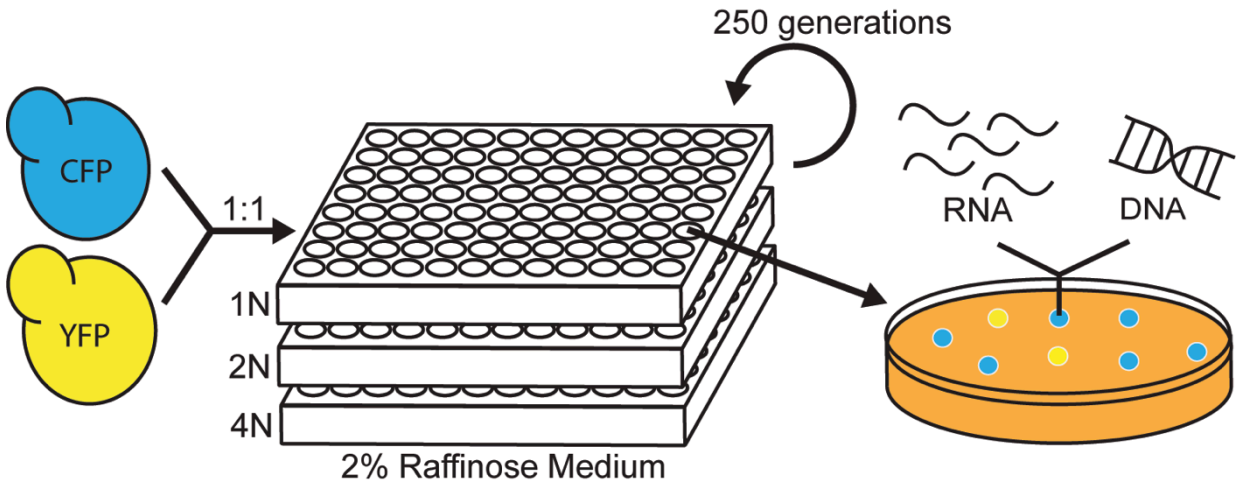


Figure 2-1 Graphical summary of the design of the experimental evolution study. Isogenic YPF or CFP expressing yeast were seeded in equal molar concentrations into 96-well plates containing 2% raffinose media. 1N, 2N, and 4N populations were evolved in raffinose for 250 generations. Evolution was monitored daily by measuring the percentage of YFP and CFP expressing cells in the population. After 250 generations clones were isolated from the populations. Additionally, RNA and DNA were collected from the isolated clones for further analysis.

2.2 Results and Discussion

2.2.1 Whole-chromosomal aneuploidy is common in tetraploid lineages

Newly formed tetraploid yeast exhibit genome instability (Storchová et al. 2006) resulting in aneuploidy, and often converging on diploidy over time in a variety of growth conditions (Gerstein et al. 2006; Voordeckers et al. 2015). To test if the clones isolated from the evolved population had experienced chromosome loss or gain we used propidium iodide straining to measure DNA content (mean G1 fluorescence) by flow cytometry in 88 clones isolated from 1Ne and 2Ne populations and 176 clones isolated from 4Ne populations. Many 4N clones underwent large shifts in ploidy; however, no aneuploidy was detected in the 1Ne or 2Ne clones (Figure 2-2).

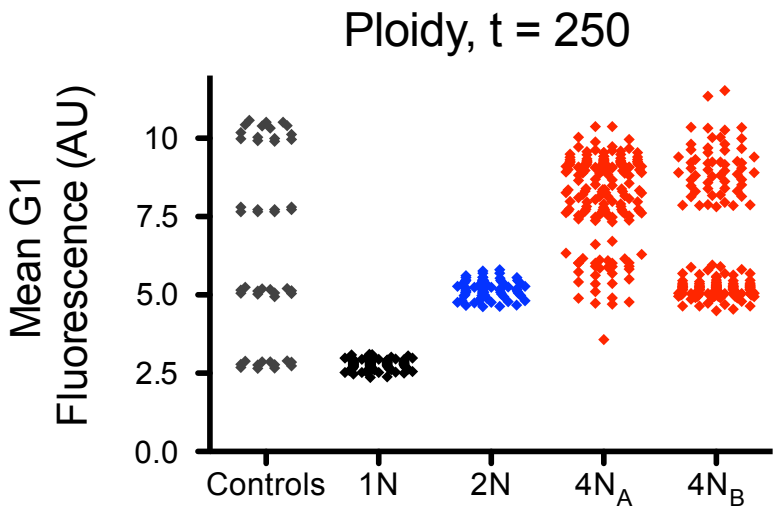


Figure 2-2 4Ne clones are highly aneuploid after 250 generations in raffinose. DNA content (1C–4C) of evolved clones at generation 250, measured as the mean G1 propidium iodide fluorescence for each evolved clone isolated from haploid (black), diploid (blue), and tetraploid (red) populations (n=88 haploid and diploid evolved clones, n=176 tetraploid evolved clones) in arbitrary units (a.u.). For reference, the DNA content of ancestral, control strains (1N, 2N, 3N, and 4N) is shown in grey.

To confirm the DNA staining and flow cytometry data, we used a combination of WGS and aCGH to determine chromosome copy number in the 1Ne, 2Ne, and 4Ne clones (Figure 2-3). In all tetraploids tested (n=30), only 1 (4Ne clone 337) remains completely 4N (Figure 2-3). For a complete list of chromosome copy numbers see appendix (Section 10). We also found that these chromosome-level alterations arose early in the evolution experiment; aneuploidy was detected in 4Ne clones isolated from generations 35 and 55 (Figure 2-4). Additionally, aneuploidy persisted as far as 500 generations (Figure 2-4). Thus, aneuploidy in the 4Ne populations is an early mechanism to generate genomic diversity in the population that may lead to adaptation.

To see if there was any trend in which chromosomes were aneuploid in the tetraploid clones, we performed pairwise correlations between each chromosome (Figure 2-5A). Pairwise patterns of chromosome copy number alterations were observed, indicating that there is a strong copy number relationship between certain pairs of chromosomes (Figure 2-5B). One potential explanation for the relationship between paired chromosomes is maintenance of stoichiometry of protein complex constituents. For example, chromosome VI disomes (1N+chromosome VI) are inviable due to an imbalance between *TUB1* encoded on chromosome XIII and *TUB2* encoded on chromosome VI (Torres et al. 2007; Anders et al. 2009). Thus, the pairs of chromosome segregating together in the aneuploid strains may encode proteins in which stoichiometry is important. Alternatively, pairs of chromosomes may encode proteins that, when overexpressed together, are beneficial to growth in raffinose.

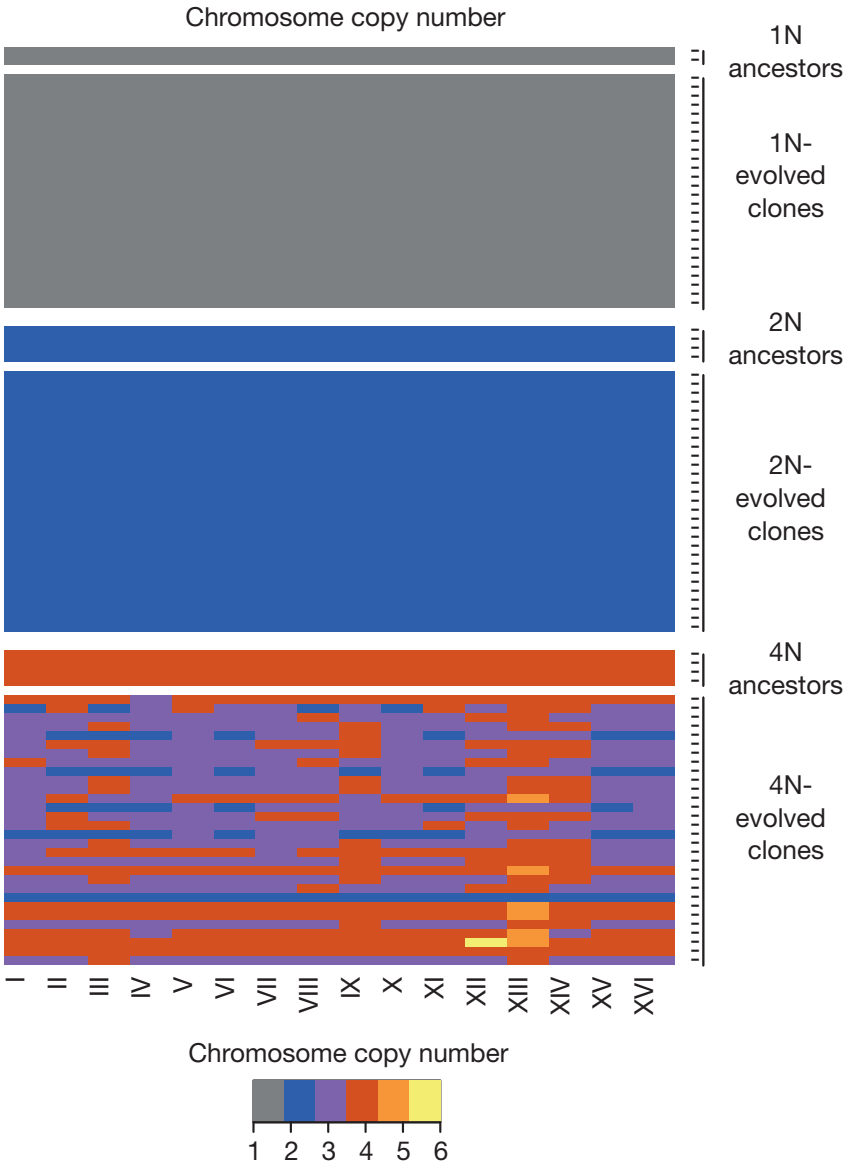


Figure 2-3 Chromosome copy number in evolved clones. Heat map of chromosome copy number data obtained from aCGH and WGS for the ancestral and evolved 1N, 2N, and 4N clones at generation 250; color key at bottom. For a complete list of chromosome copy numbers see appendix (Chapter 10).

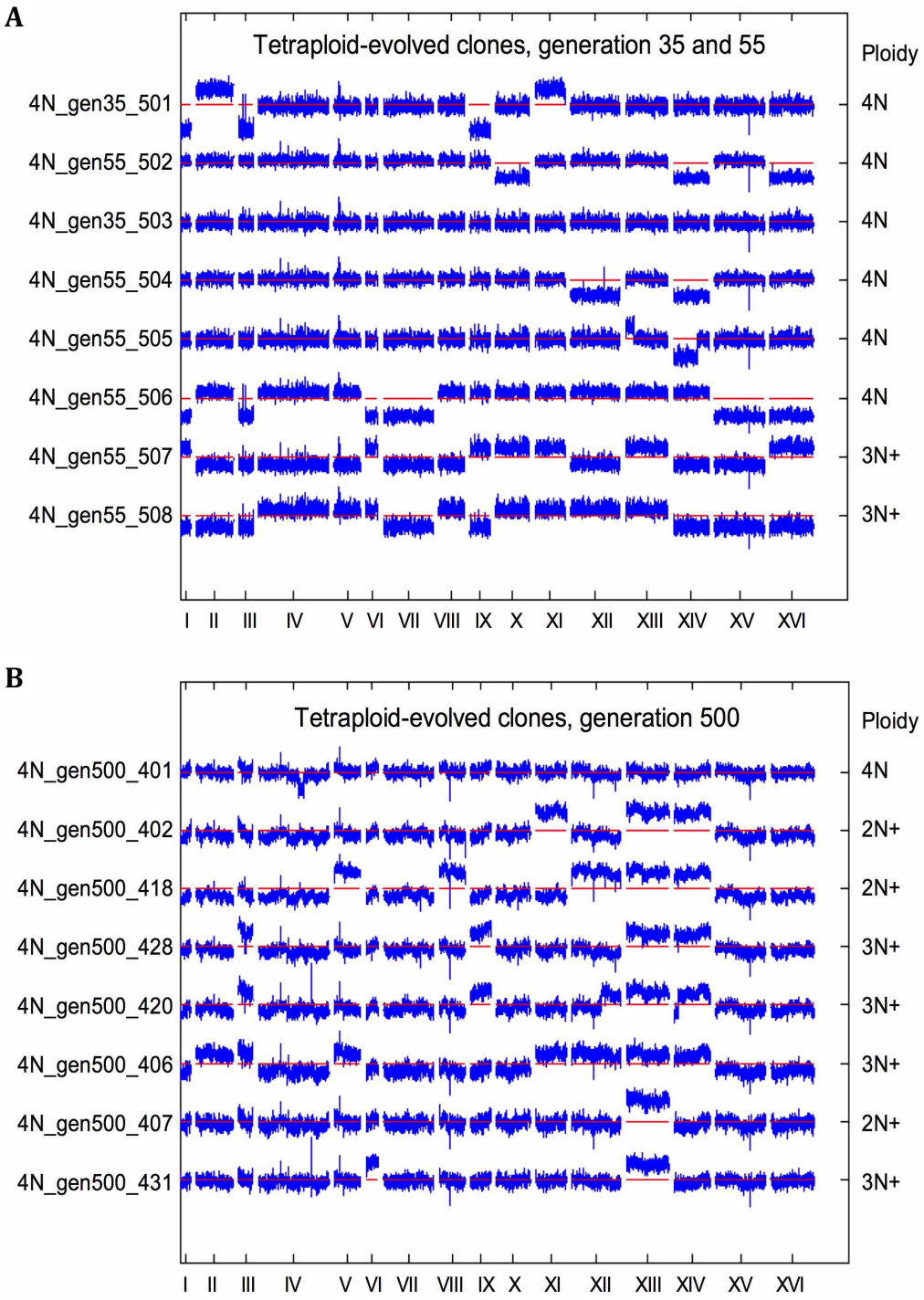


Figure 2-4 Chromosome aneuploidy in early and late generations All 4N-evolved clones at (A) generations 35 and 55 and (B) generation 500 were aneuploid for multiple chromosomes or carried large segmental chromosome aneuploidies, except for clone 4N_gen35_503, which remained tetraploid. Ploidy of the evolved clone, determined by flow cytometry, is indicated on the right, with +/- indicating chromosome aneuploidy. Figure and legend from (Selmecki et al. 2015).

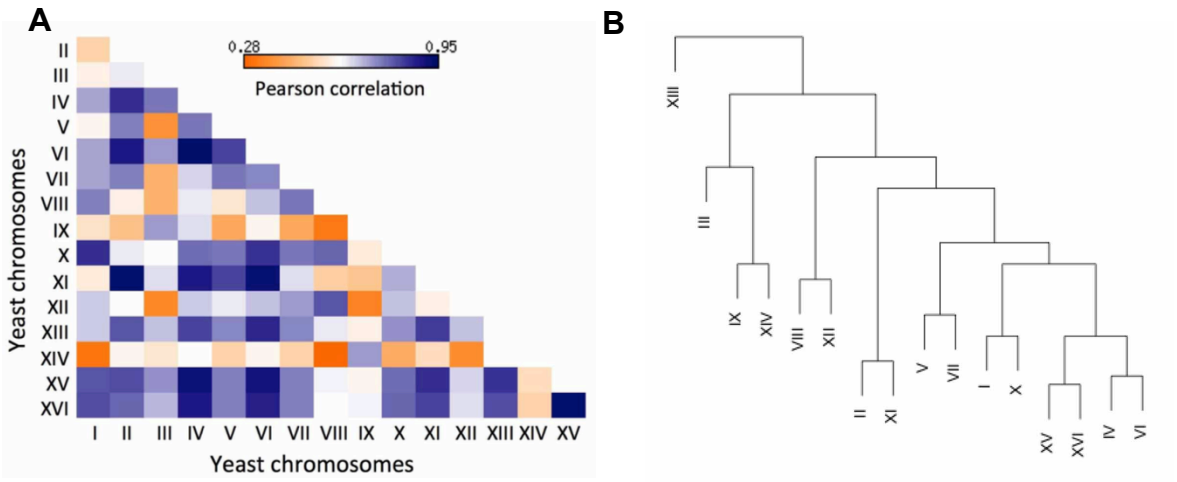


Figure 2-5 Pairwise patterns of chromosome aneuploidy A) The pairwise patterns (Pearson correlation) of all chromosome copy number alterations in the 4N-evolved clones at generation 250 (n=30, Table 10-3). The copy numbers of some chromosomes were correlated (for example, chromosome XV and chromosome XVI), whereas others were anti-correlated (for example, chromosome VIII and chromosome IX), possibly reflecting the need for gene expression balance. B) Hierarchical clustering showing the copy number relationship among the chromosomes. Figure and legend adapted from (Selmecki et al. 2015).

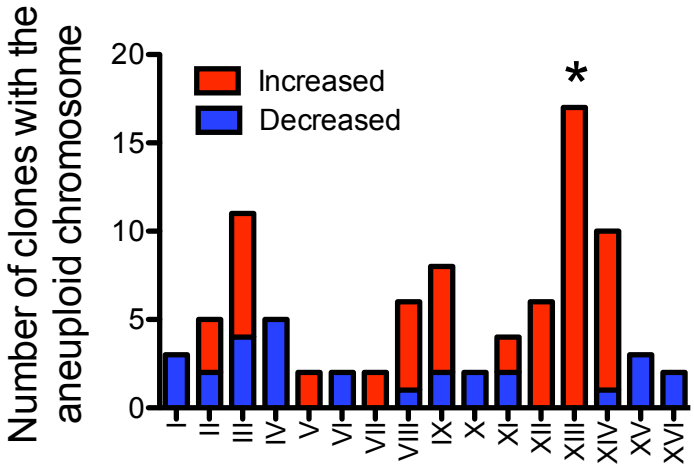


Figure 2-6 Significant enrichment of chromosome XIII aneuploidy. The number of clones with the indicated aneuploid chromosome (X-axis). The color indicates whether the copy number increased (red) or decreased (blue). The copy number of chromosome XIII (asterisk) in the 4N-evolved clones at generation 250 was significantly different from that of all other aneuploid chromosomes (Cochran–Armitage test, $p < 1x10^{-7}$).

Notably, increased copy number of chromosome XIII was significantly more common than all other aneuploidies (Figure 2-6; Cochran–Armitage test, $p < 1x10^{-7}$). Aneuploidy itself has been shown to be beneficial in certain environments: isochromosome 5 in *C. albicans* is beneficial under anti-fungal stress and chromosome XV in *S. cerevisiae* is beneficial under proteotoxic stress (Selmecki et al. 2006; Chen et al. 2012). Other examples of the advantage of specific aneuploidies in yeast has been previously reviewed (Mulla et al. 2014). This data suggests that chromosome XIII aneuploidy may be beneficial to growth in raffinose media. Chromosome XIII aneuploidy is characterized in Chapter 4.2.7.

2.2.2 Copy number variation increases with ploidy level

Copy number variation, defined here as discreet genome amplification and deletion events and segmental chromosome aneuploidies, is an important driver of evolution and individual variation (Koszul et al. 2004; Gazave et al. 2011; Iskow et al. 2012). For example CNV is common in yeast during adaptation to glucose-, nitrogen-, and sulfate- limitation and antifungal drug stress (Brown et al. 1998; Dunham et al. 2002; Selmecki et al. 2006; Gresham et al. 2008; Rancati et al. 2008; Hong and Gresham 2014). It has also been suggested that ploidy level may affect the number and repertoire of CNVs gained in adaptation: diploid yeast primarily adapted to growth in glucose-limited conditions though CNV and diploid cells gained a more diverse set of CNVs than isogenic haploids in a study designed to measure chromosome-scale mutations in yeast (Dunham et al. 2002; Zhang et al. 2013; Sellis et al. 2016). Furthermore, CNVs are often thought to be the first adaptive mutations gained in diploid and tetraploid lineages (Selmecki et al. 2015; Sellis et al. 2016).

To understand how ploidy level affected the number and type of chromosome-scale CNVs gained in adaptation to raffinose, we used a combination of aCGH and WGS to quantify CNVs in the evolved clones. After 250 generations, a segmental duplication was detected on the right arm of chromosome V in the 2Ne clone 208 (Table 10-2) and the left arm of chromosome XV in 4Ne clone 314 (Table 10-3). We detected segmental deletions on the right arm of chromosome IV in the 4Ne clones 301 and 335 (Table 10-3). These CNVs occurred at or near Ty elements. Increase in chromosome-scale CNVs in the 4Ne clones may be due to increased transposable element activity in newly formed tetraploids (Wendel 2000). Notably, we did not detect chromosome-scale CNVs in the evolved haploid strains (Table 10-1), consistent with the finding by Sellis et al. (2016).

CNV analysis indicated that many of the evolved clones had a tandem amplification of the *HXT6* and *HXT7* (*HXT6/7*^{amp}) genomic region on chromosome IV (chromosome IV: 1,154,211-1,161,315). *HXT6/7* copy number was quantified in each of the evolved clones using quantitative PCR of genomic DNA or aCGH. We found that the haploid, diploid and tetraploid strains gained amplifications in 1/24, 5/24, and 9/28 evolved clones, respectively. *HXT6* and *HXT7* are 99.7% identical and located adjacent to each other in the genome. Amplifications of *HXT6/7* are thought to occur as a result from ectopic gene conversion, which can result in extra chromosomal circular DNA, or unequal crossover, both of which occur during homologous recombination (Brown et al. 1998; Møller et al. 2015). Thus, the increased frequency of *HXT6/7*^{amp} with increasing ploidy level is consistent with increasing amounts of homologous recombination in diploids and tetraploids (Storchová et al. 2006). These data indicate that ploidy level has

a significant effect on the frequency of both chromosome-scale copy number variations as well as small-scale duplications, both of which play a role in the evolutionary outcomes of the lineages.

2.2.3 Number of mutations per strain increases with increasing ploidy

To determine the average number of mutations per strain, WGS with variant calling was performed for each evolved clone (Table 9-1). For each clone we quantified the total number of mutations, including SNVs and CNVs per strain (1Ne, Table 11-1; 2Ne, Table 11-2; 4Ne, Table 11-3). The number of mutations per strain increases with increasing ploidy level, as the 1Ne, 2Ne, and 4Ne clones had on average 2.25, 3.1, and 4.8 mutations per clone, respectively (Figure 2-7A). Given the previously measured mutation rate in yeast of 0.004 mutations per cell division per haploid genome (Lynch et al. 2008), we expect ~1, ~2, and ~4 mutations per strain after 250 generation in the haploids, diploids, and tetraploids, respectively. We recovered a greater number of mutations per clone than expected, however the number of mutations per clone is consistent with other studies in yeast grown under selection for 250 generations (Lynch et al. 2008; Araya et al. 2010; Kvitek and Sherlock 2011; Hong and Gresham 2014).There are significantly more mutations per clone in the 4Ne clone than the 1Ne and 2Ne clones ($p=7.8x10^{-6}$ and $p=2.1x10^{-3}$, respectively, paired student t-test). This result is consistent with theoretical models that predict an increased number of mutations with the increased DNA target size at higher ploidy levels (Otto and Whitton 2000). Additionally, increased DNA copy number allows for the buffering of mutations that would be deleterious in a haploid background (Korona 1999; Thompson et al. 2006; Otto 2007).

When the number of mutations per strain is adjusted for haploid genome content, there are, on average, significantly fewer mutations in the 2Ne and 4Ne clones (p=0.044 and p=0.030, respectively, paired student t-test) compared to the 1Ne clones (Figure 2-7B). One possible explanation for the decreased number of mutations per haploid genome in the 2Ne and 4Ne clones is elimination of neutral and deleterious mutations through either gene conversion events or chromosome loss. Elimination of deleterious mutations supports the advantage of yeast with higher ploidy levels in this environment.

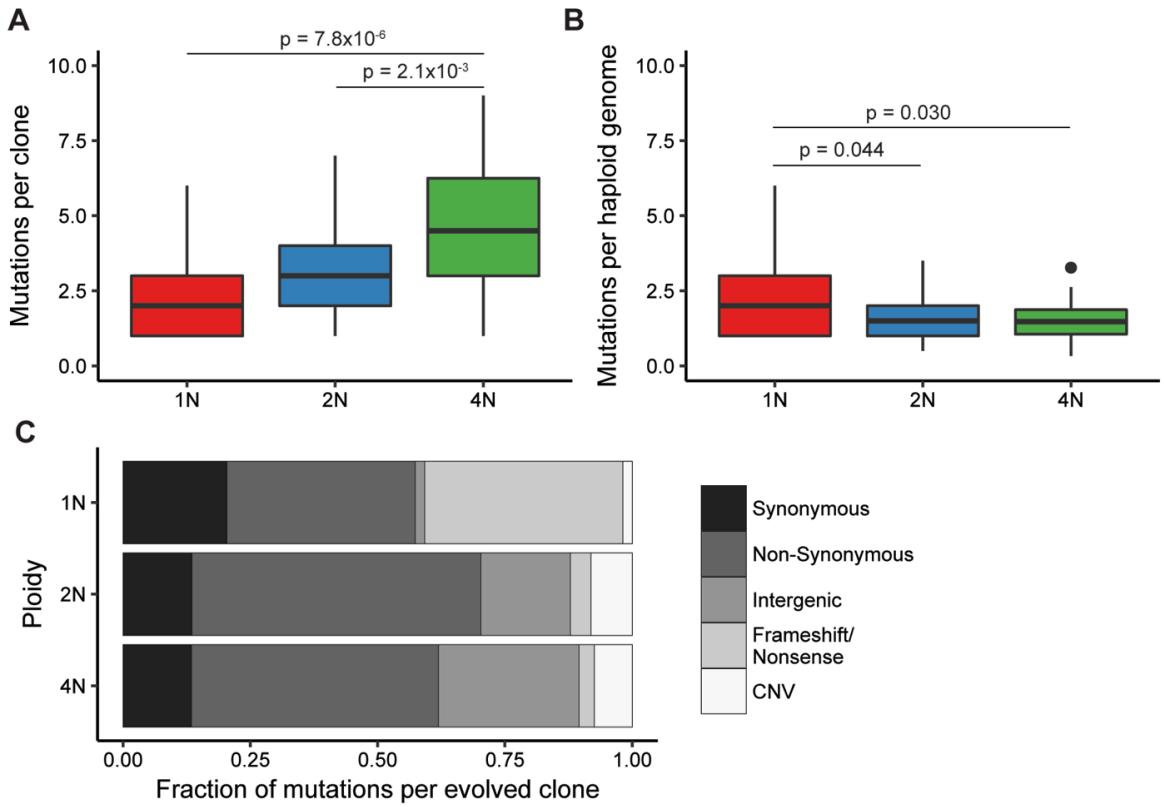


Figure 2-7 The number and spectrum of mutations per evolved clone varies by the initial ploidy. A) The average number of mutations per evolved clone derived from the 1N (red), 2N (blue), and 4N (green) evolved populations. The number of mutations includes SNVs and CNVs (including both HXT6/7amp and segmental duplications). B) The average number of mutations per clones per haploid genome in the 1N (red), 2N (blue) and 4N (green) evolved clones. Haploid genome content for each tetraploid evolved clone was by mean G1 fluorescence after 250 generations. C) The fraction of each mutation type in the 1N, 2N, and 4N experimentally evolved clones. Mutations were categorized as synonymous, non-synonymous, intergenic, frameshift or nonsense, and copy number variation (CNV). Figure and legend from (Scott et al. 2017).

2.2.4 Haploid lineages gain distinctly different mutations than higher ploidy lineages

Each genomic variant that we detected in the evolved clones was annotated as intergenic, synonymous, non-synonymous, frameshift, or nonsense (Figure 2-7C; Chapter 11). For the non-synonymous and nonsense mutations, we further identified the amino acid that was altered in the resultant protein (Table 9-1). We do not detect a difference in the average number of synonymous mutations between the 1Ne, 2Ne, and 4Ne clones (Figure 2-8). There are a greater number of non-synonymous mutations in the 2Ne (1.75×10^{-3}, paired student t-test) and 4Ne (2.40×10^{-5}, paired student t-test) clones compared to the 1Ne clones (Figure 2-8). However, the number of non-synonymous mutations per ploidy level (27/38, 45/55, 69/97 in 1Ne, 2Ne, and 4Ne clones respectively) is not significantly different from what we would expect given that ~79% of mutations that occur in the coding region are expected to be non-synonymous ($p=0.233$, $p=0.74$, $p=1$ in the 1Ne, 2Ne, and 4Ne clones respectively, exact binomial test) (Wenger et al. 2011). Therefore, the increase in non-synonymous mutations is due to the increasing number of mutations in the evolved diploid and tetraploid strains.

There are a greater number of intergenic mutations per clone in the 2Ne ($p=2.98 \times 10^{-3}$, paired student t-test) and 4Ne ($p=5.71 \times 10^{-5}$, paired student t-test) clones compared to the 1Ne clones (Figure 2-8). Given that 72% of the yeast genome is coding, the number of intergenic mutations in the 4Ne clones is greater than expected 4Ne (37/124, $p=0.0014$, exact binomial test). The greater amount of DNA in the higher ploidy strains may result in a larger number of hitchhiker mutations that do not have a fitness effect (Lang et al. 2013). However, there is one 4Ne clone, 306, which has only 3 mutations, each occurring in intergenic regions. This strain raises the possibility that

intergenic mutations may not all be hitchhikers, but rather under positive selection. It has been shown that intergenic mutations can have positive fitness effects, such as mutations in *cis*-regulatory regions (Wray 2007; Payen et al. 2016). In fact, the vast majority of human disease associated variants do not occur in the coding regions of genes (Maurano et al. 2012). Gene regulators have been preferentially maintained in the genomes of many plants, animals, and fungi after whole genome duplication, reiterating the importance of non-genic regions during adaptive evolution (Van de Peer et al. 2009).

The 1Ne clones, on average, had more frameshift and nonsense mutations per strain than the 2Ne and 4Ne clones ($p=1.77x10^{-5}$ and $9.49x10^{-6}$, respectively, paired student t-test) (Figure 2-8). Given that the vast majority of frameshift and nonsense mutations result in loss-of-function (LOF) of the gene and that LOF mutations are almost exclusively recessive, haploids are more likely to carry LOF mutations than diploids or tetraploids (Kvitek and Sherlock 2013; Lang et al. 2013; Payen et al. 2016). Additionally, the number of CNVs was greater in the 2Ne ($p=0.044$, t-test) and 4Ne clones ($p=8.70x10^{-3}$, paired student t-test). This is consistent with increasing levels of homologous recombination activity in diploids and tetraploids (Storchová et al. 2006). Thus specific properties of cells at different ploidy levels have an impact on the spectrum of mutations gained during adaptation.

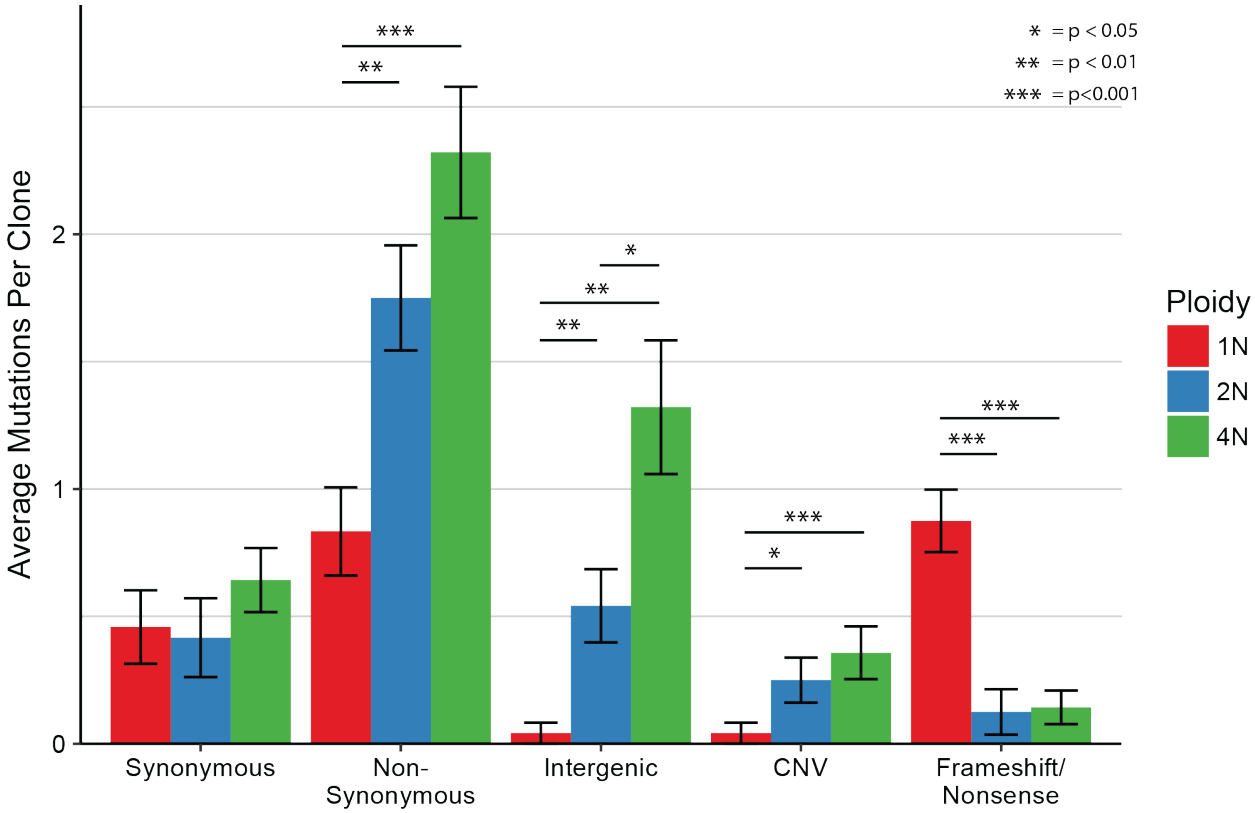


Figure 2-8 The average number of each mutation type differs by ploidy in the evolved clones. The number of synonymous, non-synonymous, intergenic, frameshift or nonsense, and copy number variation (CNV) mutations was quantified for each ploidy level. The average number of each mutation type was plotted for clones derived from 1N (red, n=24), 2N (blue, n=24), and 4N (green, n=28) experimentally evolved populations. It should be noted that the total number of mutations (shown here) includes both adaptive and neutral (hitchhiker) mutations observed within these populations. 2N and 4N clones had significantly more non-synonymous ($p=1.75x10^{-3}$, $2.40x10^{-5}$), intergenic ($p=2.9x10^{-3}$, $5.71x10^{-5}$), and CNV ($p=0.044$,$8.70x10^{-3}$) mutations per clone and fewer frameshift/nonsense ($p=1.77x10^{-5}$, $9.49x10^{-6}$) mutations per clone than 1N evolved clones. Additionally, 4N evolved clones had significantly more intergenic mutations than the 2N evolved clones ($p=0.0144$). Error bars represent standard error of the mean (SEM). All comparisons are paired student t-test. Figure and legend from (Scott et al. 2017).

2.3 Materials and Methods

2.3.1 Strains utilized in these studies

Anna Selmecki constructed the ancestral strains utilized in the experimental evolution experiments and performed the experimental evolution study. For completeness, a detailed description of the ancestral strain construction can be found in Chapter 8.1. Additionally, a detailed description of the evolution experiment can be found in Chapter 8.3.

The strains utilized in this chapter were isolated from populations at generations 35, 55, 250, and 500. Methods utilized to isolate clones can be found in Chapter 8.3. A list of strains that were sequenced for this study, and the mutations identified in these strains, can be found in Table 9-1.

2.3.2 Flow cytometry analysis of DNA content

Anna Selmecki quantified DNA content by mean G1 fluorescence of propidiom iodide staining. Briefly, cells were prepared for propidium iodide staining with modifications to optimize preparation of samples in 96-well plates. Thirty thousand cells were analyzed using a BD LSRII High Throughput Sampler. FlowJo cell cycle analysis used the Dean–Jett–Fox model to estimate the mean G1 and G2 fluorescence peaks of each strain. Control ancestral 1N, 2N, 3N, and 4N strains were analyzed in triplicate with the evolved strains.

2.3.3 Illumina sequencing

Library preparation was performed primarily by Amber Scott with assistance by Anna Selmecki. Read mapping and variant calling was performed by Phillip Richmond

and Robert Thomas. Amber Scott and Anna Selmecki performed visual inspection of all variant calls.

In line barcoded library preparation. Clones selected for WGS were cultured overnight from -80°C stocks in 2 ml Synthetic Complete + 2% raffinose medium. Genomic DNA was isolated using phenol–chloroform–isoamylalcohol (24:25:1) and bead beating. Libraries were prepared as described (Hittinger et al. 2010). Briefly, DNA was sheered with Diagenode Bioruptor (UCD-200) to a median size of 300–500 bp, end-repair was performed with an NEB Next End repair kit (NEB E6050L), and fragments were A-tailed with Klenow fragment (NEB M0212L). Custom adaptors with in-line barcodes were ligated overnight. Adaptor ligated fragments were size selected on 1% TBE agarose gel stained with Sybr Gold (Invitrogen S-11494) for fragments between 400 and 600 bp and isolated using Qiagen Gel Extraction Kit (28706). Libraries were amplified for 12 cycles with Illumina PE PCR primers 1.0 and 2.0. Libraries were pooled and underwent additional size selection for fragments of 400–600 bp.

Following the Selmecki et. al. publication, two additional evolved clones were sequenced by Anna Selmecki and Robert Thomas (1Ne-131 and 1Ne-132). Libraries were prepared using the NexteraXT DNA Sample Preparation Kit with NexteraXT index kit, and following the manufacturer's instructions (Illumina). Libraries were sequenced using an Illumina MiSeq (Creighton University).

Raw data. The genomes were sequenced on an Illumina HiSeq 2500 at the University of Colorado at Denver Next Generation Sequencing Facility or MiSeq at Creighton University. The data, which had an inline barcode, were de-multiplexed by the sequencing facility into individual sample R1/R2 files—one file for each read in the pair.

The barcodes were removed before mapping using Fastx_trimmer v0.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit/). Read trimming from the 5' end of the R2 reads was performed on a sample-specific manner trimming anywhere from 0 to 28 base pairs using an in-house script and Fastx_trimmer.

Mapping. Reads were mapped to the *S. cerevisiae* reference sequence for the laboratory yeast strain S288c reference genome (*S. cerevisiae* genome obtained 28 July 2010 from the Saccharomyces Genome Database FTP site: http://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/; actual genome: http://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/S288C_reference_genome_R63-1-1_20100105.tgz). The reads were mapped using the Bowtie2 v2.0.2 (Langmead and Salzberg 2012) local alignment strategy, allowing for multiple mapping, and setting the following options:--very-sensitive-local -I 180 -X 1,000–score-min G,70,8. The mapped reads then underwent file format conversion into the binary format for downstream analysis using Samtools view, sort, and index v0.1.18 (Li et al. 2009). Post-alignment to the genome, duplicate pairs resulting from PCR over-amplification were removed using Samtools rmdup, eliminating 1–5% of the paired reads. The reads were re-aligned over potential indel sites using the Genome Analysis Toolkit RealignerTargetCreator and IndelRealigner v2.4-9 (McKenna et al. 2010; DePristo et al. 2011).

Variant calling and refinement. Variant calling was performed on the tailored read mappings using Genome Analysis Toolkit UnifiedGenotyper v2.4-9 (McKenna et al. 2010; DePristo et al. 2011). For the haploids and diploids, SNPs were called using default parameters; for the higher ploidy strains, the ploidy option was increased to 5n,

which allowed identification of mutations at allelic frequencies down to 5% alternative allele representation. Variant lists were combined on the basis of ploidy-type using Genome Analysis Toolkit CombineVariants (McKenna et al. 2010; DePristo et al. 2011). SNPs and short indels were compared with the ancestral set of mutations using in-house scripts to generate a set of non-ancestral mutations. These mutations were filtered for alternative allele support and allelic frequency (more than two reads supporting an alternative allele for coverage 10x to 20x, and more than four reads supporting an alternative allele for coverage more than 20x). The filtered mutations were manually inspected using the Integrative Genome Viewer v2.1.19 (Thorvaldsdóttir et al. 2013) to refine the set and remove mapping artifacts such as strand representation bias, regional mapping quality issues from non-unique mapping, and artifacts of homopolymer and simple repeat alignments. We then Sanger sequenced variants with low read support (fewer than five reads supporting an alternative allele), as well as a subset of the other medium- and high-confidence variants. The final set of evolved variants discovered was annotated (synonymous, nonsynonmous, frameshift, etc) relative to the yeast gene annotations using an in-house script.

2.3.4 Sequencing quality assessment

Sequencing quality assessment was performed by Phillip Richmond, Amber Scott, and Anna Selmecki. Because our sequencing was highly multiplexed, quality assessment on the sequencing data was necessary to eliminate strains without adequate genome coverage. For the haploids and diploids, we determined the necessary depth of coverage to recover mutations in two ways. First, we took the set of ‘strain-background’ mutations, which were identified by filtering the ancestral variant

calls for a conservative, high-quality (quality = 100), homozygous set of locations. Each strain was then queried for its ability to recapitulate these variants, reporting a percentage overlap between each strain's variant calls and the set of background variants. Any strain with less than 97% of the background mutations was dropped from further consideration. Additionally, we examined the impact of sub-sampling down to various depths to investigate the impact of lower coverage on recovering variants. This was done using Picard's DownsampleSam.jar (version 1.72, http://broadinstitute.github.io/picard/) on two high-coverage diploid strains to randomly down-sample the coverage to 100x, 50x, 25x, and 10x coverage. We examined the down-sampled data sets and found that the strain-unique SNPs could be captured even at a level of 10x coverage. Using this information, we set minimum coverage requirements for each strain on a genome-wide scale to eliminate strains without adequate genomic representation. We analyzed depth of coverage on all of the mapped data with BEDTools genomeCoverageBed v2.16.2 (Quinlan and Hall 2010). The per-base coverage was then analyzed using an in-house script to produce statistics on minimum coverage per allele, average coverage, etc.

2.3.5 SOLiD Sequencing

Phillip Richmond and Anna Selmecki prepared the SOLiD sequencing libraries and Phillip Richmond performed all sequence analysis for the SOLiD sequencing libraries.

Library preparation. We performed a pilot experiment on seven strains using SOLiD paired-end sequencing (Table 9-1). Clones selected for SOLiD sequencing were cultured overnight from -80°C stocks in 4ml Synthetic Complete + 2% raffinose medium. Genomic DNA was isolated using QIAGEN Genomic-Tip 1000 according to the

manufacturer's instructions. SOLiD library preparation and sequencing was performed by the Molecular Biology Core Facility at the Dana-Farber Cancer Institute, according to the manufacturer's instructions (Applied Biosystems, Life Technologies).

Mapping. The sequencing reads were mapped to the *S. cerevisiae* reference genome (see 'Mapping' in the Illumina section above) using multiple different mapping software including BWA v.0.5.9 (Li and Durbin 2010), NovoAlignCS v1.01.05 (Novocraft), Bfast v0.6.5a (Homer et al. 2009), and BowtieCS v0.12.7 (Langmead et al. 2009). BowtieCS and BWA were used in the downstream variant calling and copy number changes, while NovoAlignCS and Bfast served as added support in manual inspection of variants. After mapping, the reads were post-processed for local realignment using SRMA v0.1.15 (Homer and Nelson 2010) and Samtools BAQ v0.1.18 (Li et al. 2009).

Variant calling and refinement. SNPs, small insertions and deletions (indels) were called from the post-processed reads using Samtools Mpileup v0.1.18 (Li et al. 2009), VARiD v1.0.7f (Dalca et al. 2010), and Freebayes (v0.8.9, http://bioinformatics.bc. edu/marthlab/FreeBayes). Samtools and VARiD variant calls were used to identify the strain background (ancestral variants relative to the reference). These variants were filtered on the basis of reads supporting the allele in both directions, quality score of the call, and adequate read coverage over the call. Once filtered, all of the variant calls for the evolved strains were merged and compared with the ancestors. Variations were verified by manual inspection followed by Sanger validation for both a set of randomly sampled loci and regions of disagreement between different combinations of the mapping software and the variant callers (that is,

dinucleotide SNPs and multiple indels within a single read). The resulting set was later used for identification of strain-unique variants in the evolved strains.

To identify strain-unique variants, Freebayes, a variant caller capable of higher-ploidy (ploidy = 2N), was used. Freebayes has the ability to set the assumed ploidy over a genomic region to adjust the expected distribution for allelic frequency. The assumed ploidy was determined using aCGH as well as copy number changes implied by shifts in depth of sequencing. The Freebayes called variants on each evolved progeny were then cross-referenced with the ancestral variants to produce strain-unique variants (Table 9-1These variants were then manually examined in Integrative Genomics Viewer (Thorvaldsdóttir et al. 2013) and validated by Sanger sequencing. PCR amplification and Sanger sequencing used primers located approximately 200 bp on either side of the sequence variants on DNA from both the evolved clone and the ancestor.

2.3.6 Chromosomal CNV identification

Phillip Richmond identified chromosomal copy number variations in the whole genome sequencing data. Identification of chromosomal CNVs used HTSeq v0.6.1 (Anders and Huber 2010) in conjunction with custom scripts. HTSeq performs coverage estimations on a per-gene basis, and the custom scripts provided normalized $\log_2$ fold change between each sample and the ancestral haploid strain. Estimates on chromosomal copy number were inferred using the median value for the $\log_2$ fold change on a chromosome-by-chromosome basis. We implemented the Cochran–Armitage test to determine whether chromosome XIII had a trend for higher copy number, relative to the copy number observed for all chromosomes in the tetraploid-evolved clones (Table 10-3, Figure 2-6). This trend analysis is similar to a chi-squared

test, but considers whether there is a significant trend or direction to the observed data set (chromosome XIII copy number).

2.3.7 Microarray aCGH

Anna Selmecki measured chromosome copy number in the ancestral and evolved clones by array CGH as follows. Fluorescently labeled DNA was prepared for CGH as described previously (Selmecki et al. 2005). Genomic DNA from all experimental strains was compared with the same pool of genomic DNA from the ancestral strain background PY3295 (BY4741, Research Genetics). Agilent yeast DNA 4344K microarrays (ChIP-on-chip Kit) were used for the hybridization according to the manufacturer's instructions (Agilent Technologies) with several modifications (M. Dunham online protocols, http:// dunham.gs.washington.edu/protocols.shtml). Briefly, 2.0 mg of HaeIII-digested (New England Biolabs) genomic DNA was labeled with 2.1 ml of Cy3 or Cy5 (CyDye–Cy3-dUTP or CyDye–Cy5–dUTP, Amersham GE Healthcare). 300ng of Cy3-labelled DNA (experimental strains) was mixed with 300ng of Cy5-labelled DNA (control DNA) and the volume was brought to 44ml with nuclease-free water. Blocking buffer and hybridization buffer 2xHiRPM (Agilent Technologies) were added, and 100ml was applied to each sub-array; the microarray was hybridized at 65°C for 17 h and then washed, scanned, and analyzed according to the manufacturer's instructions. Agilent Feature Extraction data were converted from $\log_{10}$ ratios to $\log_2$ ratios and plotted using Treeview (Saldanha 2004) and a custom Matlab script. A $\log_2$ ratio of zero (baseline) indicates no difference in DNA copy number between reference and experimental samples (Selmecki et al. 2005; Selmecki et al. 2006).

2.4 Conclusions

While the selection pressure (i.e. growth in raffinose, low constant fructose availability) are constant between the haploid, diploid and tetraploid populations in our study, the mechanisms driving evolution in the haploid cells differs from those in the diploid and tetraploid cells. Overall, both the number and spectrum of mutations differs in the 1Ne clones compared to the 2Ne and 4Ne clones. However, it is unknown to what extent the mutations acquired in these strains are beneficial mutations or hitchhikers (Lang et al. 2013). To understand the impact of these mutations on function, I utilize expression analysis in Chapter 3.

As expected by theoretical models, the number of mutations acquired per strain in the evolution study increased with increasing ploidy (Figure 2-7A)(Ohno 1970; Otto and Whitton 2000). This may suggest one mechanism by which the tetraploid clones are adapting more rapidly than haploids and diploids in the evolution study. However, the diploids, which acquire more mutations than the haploids, do not have a significantly greater rate of adaptation than the haploids (Figure 1-4C). Therefore, it may not be the number of mutations, rather the types of mutations that are critical for the increased rate of adaptation in the tetraploid lineages.

The haploid evolved clones acquire a different spectrum of mutations than the diploid or tetraploid clones (Figure 2-7C). This difference is driven primarily by a greater number of recessive mutations in the haploids, such as nonsense and frameshift mutations, which are inaccessible to the diploid and tetraploid lineages (Figure 2-8). On the other hand, diploid and tetraploid clones acquire a greater number of non-synonymous and intergenic mutations than the haploid clones (Figure 2-8). In all ploidy

levels the number of non-synonymous mutations is not significantly different than expected by the exact binomial test (Hong and Gresham 2014). However, the number of intergenic mutations in the haploid evolved clones is fewer than expected (exact binomial test, $p=8.02x10^{-7}$). While it is not clear why the haploids have fewer intergenic mutations, non-coding variants are increasingly thought to contribute to disease (Maurano et al. 2012). Thus, otherwise deleterious intergenic mutations may be buffered in higher ploidy cells while causing decreased fitness in haploid lineages (Otto and Whitton 2000).

The most striking difference between the types of mutations acquired at the different levels of ploidy is in increase in copy number variations observed with increasing ploidy. Higher ploidy strains amplified the *HXT6/7* genomic region with higher frequency and acquired segmental duplications more often (Figure 2-8). Additionally, tetraploid evolved clones became highly aneuploid, whereas haploid and diploid clones remained 1N or 2N, respectively, throughout the evolution experiment (Figure 2-3). Importantly, chromosome-scale copy number variation and aneuploidy occurred early in adaptation to raffinose (Figure 2-4A). Thus, copy number variation may be a key mechanism of early adaptation in tetraploid lineages.

3 PATHWAYS OF ADAPTATION

A portion of this chapter is currently under review and was submitted as:

Amber L Scott, Phillip A. Richmond, Robin Dowell, Anna M. Selmecki. The influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source. *Under Review.*

3.1 Introduction

In the second half of the study, I focus on the adaptive pathways utilized in the haploid, diploid, and tetraploid populations to adapt to growth in 2% raffinose media. In particular, I examine the extent of parallel evolution in the clones sequenced in the previous chapter. Parallel evolution is defined as the same change having evolved independently multiple times (Wichman 1999). This can mean parallel gene expression alterations or adaptive mutations in the same gene (Cooper et al. 2003; Fisher and Lang 2016). Experimental evolution studies in yeast have observed parallel expression changes in replicate populations (Ferea et al. 1999; Gresham et al. 2008). Additionally, studies have recovered mutations in the same genes from independent populations (Kao and Sherlock 2008; Hong and Gresham 2014; Voordeckers et al. 2015). However, these studies were performed on only a small number of replicate populations. Our study is the largest study to date to examine adaptive pathways in clones isolated from independent populations.

To determine the pathways of adaptation, I performed whole genome expression analysis or quantitative PCR in over 100 evolved clones derived from the haploid, diploid, and tetraploid populations. I demonstrate that the key adaptive mutation in the

evolved clones is predicted by a gene expression signature of just 5 genes. I further determine the primary adaptive mutation for all but 5 evolved clones and assess to what extent the different ploidy levels exhibit parallelism in the modes of adaptation. The adaptive mutations identified encompass a narrow set of genes, indicative of parallel evolution between strains at each ploidy level. However, the initial ploidy affected the relative likelihood of a given adaptive pathway.

3.2 Results and Discussion

To further understand the mechanisms by which strains of differing ploidy level adapted to growth in raffinose medium, we examined gene expression changes in the evolved clones after evolution. We utilized strand-specific RNA sequencing followed by differential expression analysis on 8 evolved clones (2 haploid, 2 diploid, and 4 tetraploid) as well as both the diploid and tetraploid ancestral strains. Furthermore, we combined gene expression data with the previously described (Chapter 2) whole genome sequencing data to correlate adaptive mutations with their gene expression phenotype.

3.2.1 *Tetraploids ancestor exhibits greater stress in raffinose*

Tetraploid yeast strains differentially express only a few genes compared to isogenic diploids when grown in rich media (Galitski et al. 1999; Storchová et al. 2006; Wu et al. 2010). However, this may not be true in other growth conditions. Thus, we initially sought to determine the relative impact of growth in raffinose on the expression profiles of the diploid and tetraploid ancestors. We performed differential expression analysis and gene ontology (GO) enrichment between the diploid and tetraploid ancestral strains grown in raffinose (Figure 3-1). There were 177 genes differentially

expressed between the diploid and tetraploid ancestors (Table 12-2). Genes over-expressed in the tetraploid ancestor are significantly enriched for processes involved in starvation, including energy reserve, glycogen, glucan, trehalose, polysaccharide and carbohydrate metabolic processes as well as other biosynthetic processes (Table 13-3). These processes indicate that the tetraploid ancestral strain may have to reach further into energy reserves than the diploid ancestor.

This differential expression between the 2N and 4N ancestors is not entirely explained by increased metabolic need. The differentially expressed genes are also enriched for genes involved in oxidative stress, suggesting that the tetraploid is more stressed than its diploid counterpart when grown in raffinose media (Table 13-3). To test if the tetraploid ancestor exhibited gene expression alterations related to the environmental stress response, we determined the enrichment of the environmental stress response (ESR) genes in tetraploid ancestral expression profile (Gasch et al. 2000). The environmental stress response (ESR) is defined in yeast as a specific set of genes that are commonly altered in expression in response to a wide range of stressors (Gasch 2002). For this analysis we utilized a gene set enrichment approach where the enrichment score is determined from the correlation of up-regulated and down-regulated gene sets, in our case ESR genes, with a rank ordered list, in our case the $\log_2$-expression ratio between the tetraploid ancestor and diploid ancestor (Subramanian et al. 2005; Lamb et al. 2006). The expression of the ESR genes is positively enriched in the tetraploid ancestor, with an enrichment score of 0.996 out of a maximum of 1, indicating that the total gene expression profile of the 4N ancestor is consistent with the environment stress response (Figure 3-2).

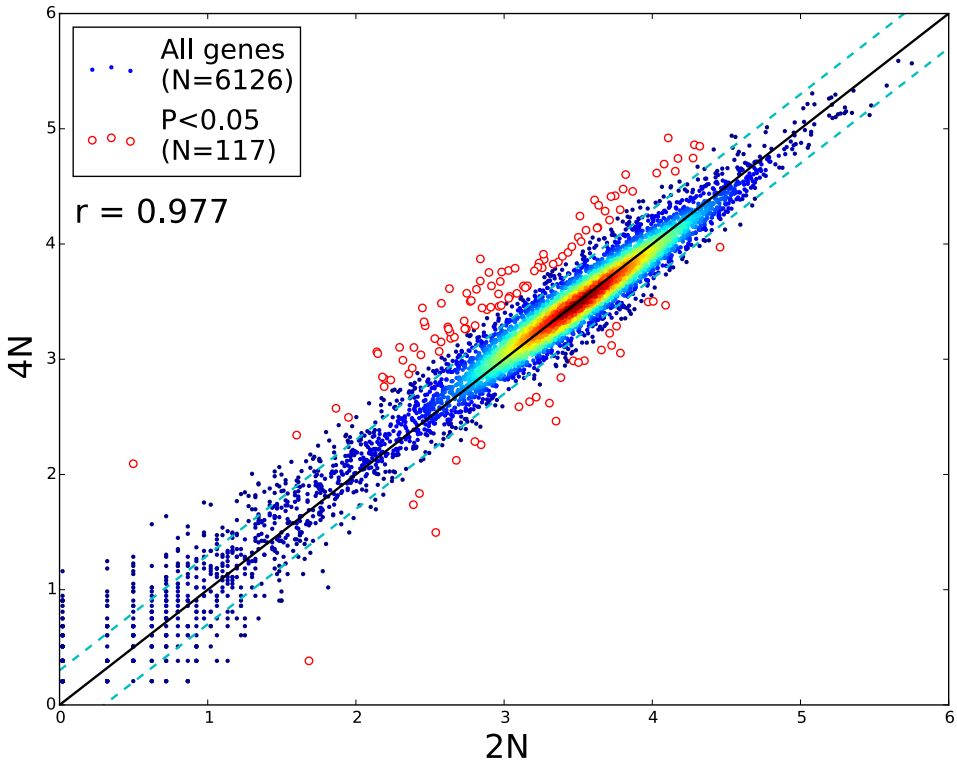


Figure 3-1 Comparative gene expression between the 2N and 4N ancestor. Normalized $\log_{10}$ gene expression in the tetraploid ancestor (Y-axis) compared to the diploid ancestor (X-axis). The gene expression values for each gene are the normalized read counts calculated by DESeq. Significantly differentially expressed genes (Adjusted P<0.05) are denoted by red circles. The density of genes is indicated by color from solid blue dots (low density) to red dots (high density). The dashed cyan lines indicate 2-fold differential expression.

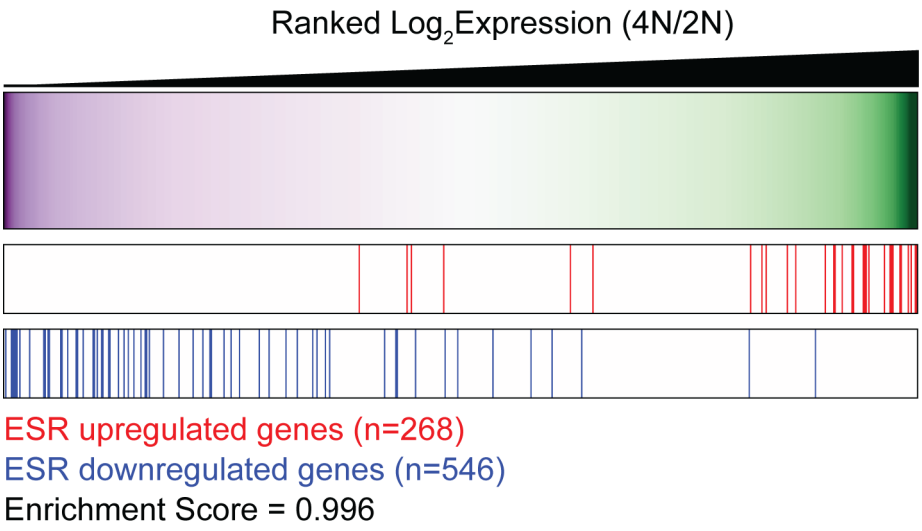


Figure 3-2 ESR enrichment in the tetraploid ancestral strain. The log_2expression ratio is ordered from lowest (down-regulated genes, purple) or highest (up-regulated genes, green). The position of known upregulated (red) or downregulated (blue) genes in the environmental stress response are labeled by lines. Enrichment scores range from +1 to -1, indicating positive or negative enrichment, respectively. An enrichment score of 0 indicates no enrichment. ESR genes are enriched in the tetraploid ancestor relative to the diploid ancestor with an enrichment score of 0.996.

To confirm that the stress response was exclusive to the tetraploid ancestor grown in raffinose, we measured the expression of five iso-enzyme pairs in the 1N, 2N, and 4N ancestral strains grown in glucose-limited media (SC + 2% raffinose). Each pair of iso-enzymes consists of two genes with similar enzymatic function; the expression of one gene in the pair is upregulated in ESR and the expression of the other gene is unaffected by ESR. Thus, the non-ESR iso-enzyme internally controls for expression changes that may be environment or ploidy specific. While the trend indicated that the 4N ancestral strain specifically over-expresses the iso-enzymes associated with ESR, the gene expression differences are not significant (Figure 3-3, paired student t-test). Stress levels have been associated with an increase in mutation rate (Bjedov et al. 2003; Galhardo et al. 2007) and genome instability (Forche et al. 2011; Chen et al. 2012). This increase in ESR in the tetraploid background may suggest the mechanism behind increased adaptation rate in the tetraploid lineages compared to haploid and diploid lineages.

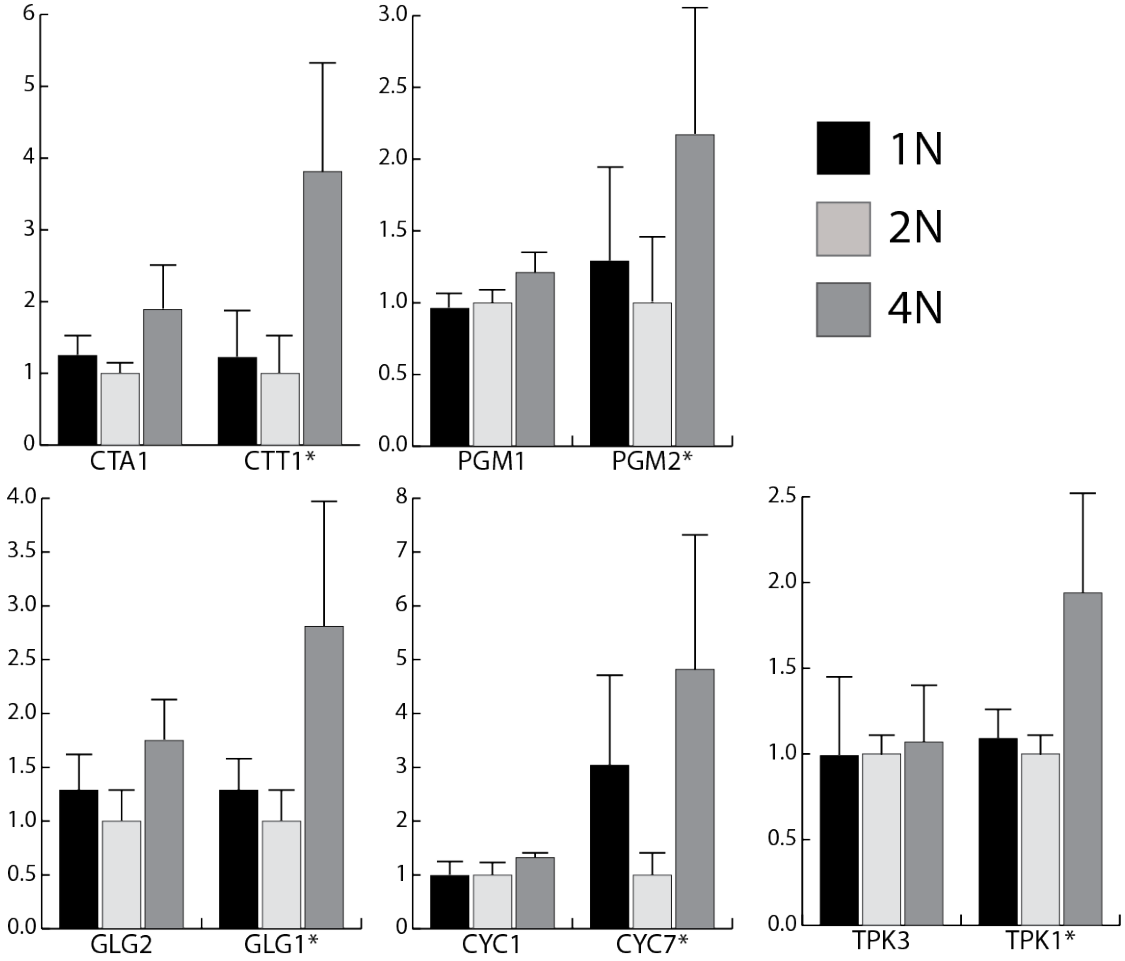


Figure 3-3 Increase in environmental stress response in 4N ancestral strain. Relative normalized gene expression for iso-enzyme gene pairs in haploid (black), diploid (light grey), and tetraploid (dark grey) ancestral strains. Gene expression is normalized internally to *ACT1* expression and relative to the diploid ancestor. The enzyme in each pair associated with the environmental stress response is indicated by an asterisk. Error bars represent the SEM, n=5.

3.2.2 Evolved clones show few expression differences from ancestral strain

To understand how the evolved clones adapted, we compared the gene expression profile when grown in raffinose of each evolved clone to the diploid ancestor (Figure 3-4, Figure 12-1, Figure 12-2). We found few alterations in gene expression in the evolved clones, on average only 26 genes were differentially expressed when compared to the diploid ancestor (Chapter 12). In contrast, previous gene expression studies on yeast strains evolved in glucose-limiting conditions found major metabolic shifts in the evolved clones (Ferea et al. 1999; Gresham et al. 2008). There are two major differences in our study compared to the previous studies that may affect the expression of metabolic genes: we propagated cells by serial transfer rather than growth in a chemostat and cultures were adapted in raffinose rather than low glucose. Given the small changes observed in expression profiles, the increased growth rate of the evolved clones in raffinose must therefore be due to only small changes in the expression of many genes, large changes in a few key genes, or due to alterations at the protein level.

3.2.3 Glucose transport is upregulated in the evolved clones

We performed gene ontology (GO) term enrichment on the genes that were differentially expressed between the diploid ancestor and each evolved clone to determine the pathways that contributed to the adaptation (Chapter 13). There are two terms commonly enriched in the evolved clones: *phosphate ion transport genes* are commonly downregulated and *carbohydrate transport genes* are frequently upregulated in the evolved clones.

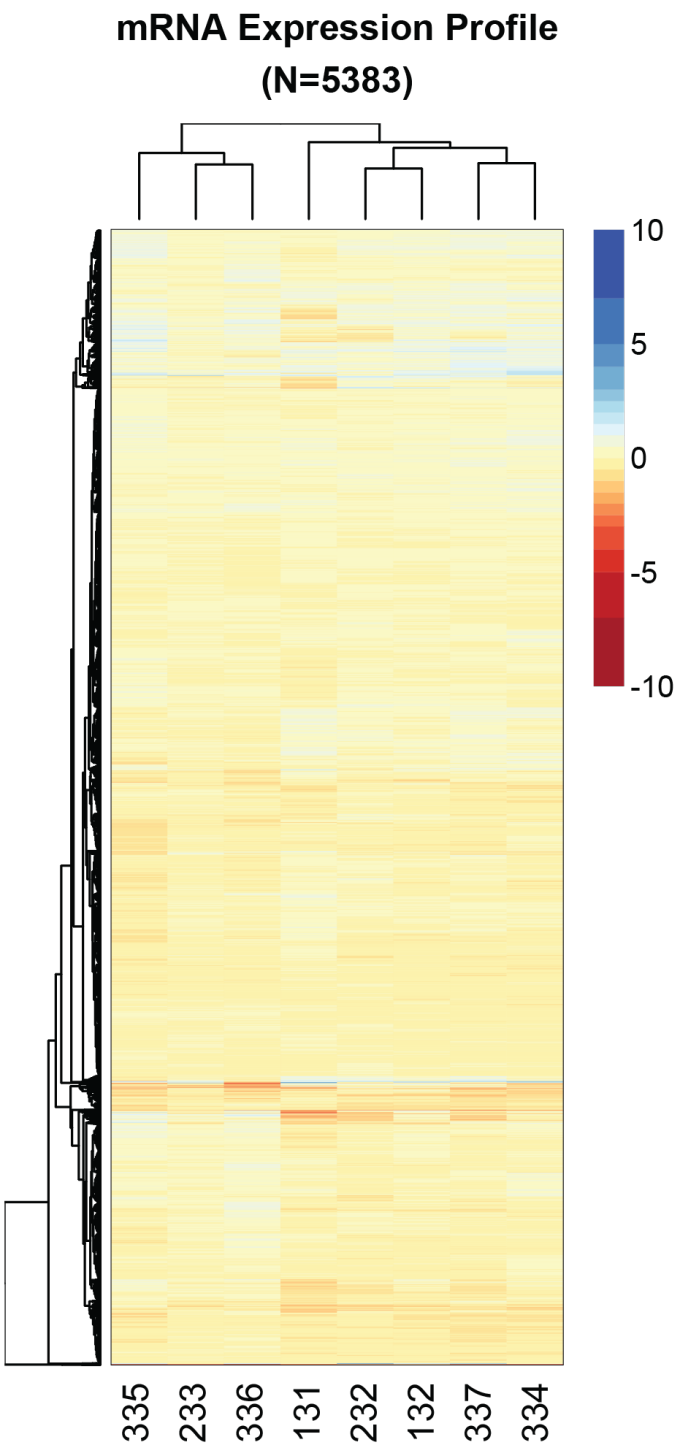


Figure 3-4 mRNA expression profile in the evolved clones Log_2-transformed gene expression ratio between evolved clones (131, 132, 232, 233, 334, 335, 336, 337) and the diploid ancestral strain. Hierarchical clustering was performed for both genes (y-axis, n=5383) and evolved clones (x-axis). Figure and legend from (Scott et al. 2017).

Phosphate ion transport is decreased in all of the evolved clones; primarily driven by the overexpression of *PHO84* and *PHO89* in the diploid ancestor (Figure 3-5A). We measured, by qRT-PCR, the expression of *PHO84* in the ancestral haploid and tetraploid strains. Only the diploid ancestor consistently had elevated *PHO84,* though with high variability (Figure 3-5B). Regulation of the phosphate genes is controlled by a hysteretic switch which may contribute to variability in activation of the phosphate transporters (Raser and Shea 2006).

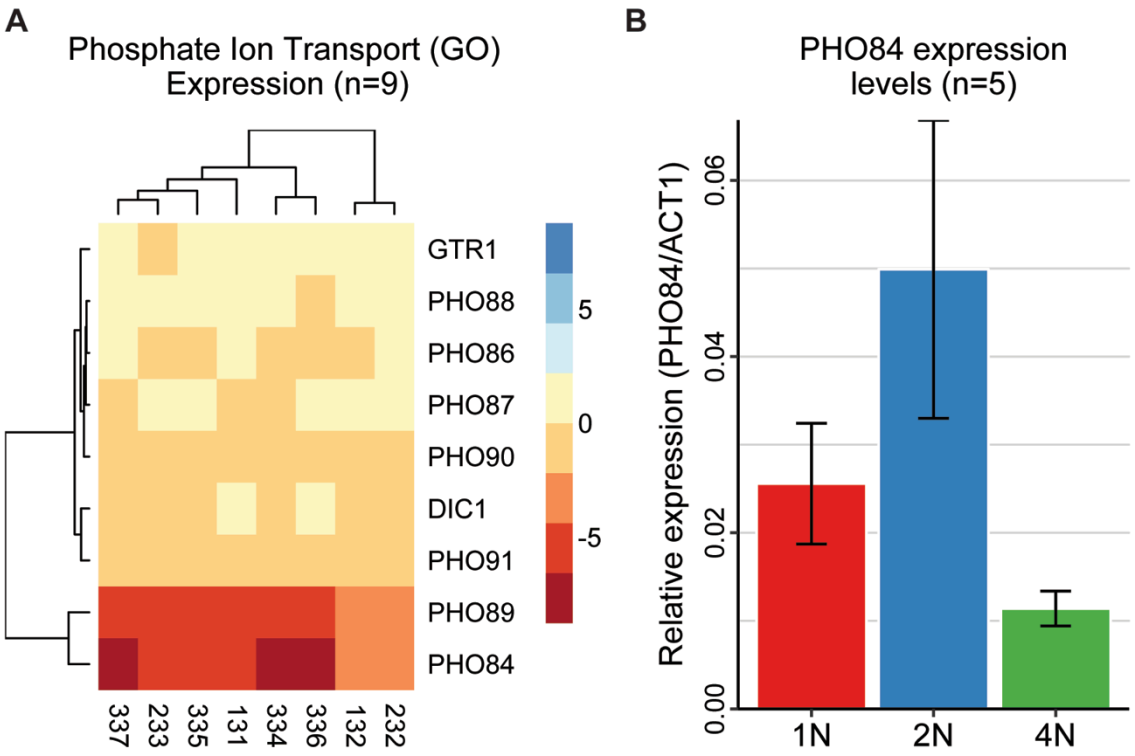


Figure 3-5 Evolved clones differentially regulate phosphate ion transport. A) Log_2-transformed expression of genes annotated with the GO-function of phosphate ion transport.. B) qRT-PCR expression of *PHO84* relative to *ACT1* expression in the haploid (red), diploid (blue), and tetraploid (green) of 5 biological replicates. Error bars represent the standard error of the mean (SEM). Figure and legend from (Scott, AL et al. 2017).

Consistent with adaptation to a carbon stress, the differentially expressed genes also show enrichment for carbohydrate transport, specifically the hexose transporters (Figure 3-6). Carbohydrate transport is a common mode of adaptation to glucose-limited growth conditions; previous studies have identified mutations in the carbohydrate sensing and transport genes in yeast grown in low-glucose environments (Brown et al. 1998; Gresham et al. 2008; Kao and Sherlock 2008; Koschwanez et al. 2013; Kvitek and Sherlock 2013). To understand how carbohydrate transport was altered in our raffinose-evolved clones, we performed hierarchical clustering of the RNA expression of genes annotated as “carbohydrate transport” in the evolved clones relative to the diploid ancestor (Figure 1-3). Of the ontological carbohydrate transport genes, only the hexose transporters are differentially expressed.

To test if ploidy level itself has an effect on the expression of the hexose transporters in different carbon sources, we measured the expression of *HXT1, HXT2, HXT3, HXT4, HXT6/7, and SUC2* in SC + 0.1% glucose, SC + 2% glucose, SC + 2% galactose, and SC + 2% raffinose (Figure 3-7). Ploidy level itself does not impact the expression level of the hexose transporters. However, the carbon source does influence the level expression of the different transporters, as previously described (Chapter 1.8). Notably, the ancestral strains over-express *HXT6/7*, *HXT2*, and *SUC2* when grown in raffinose media (Figure 3-7D). Additionally, galactose, a non-fermentable carbon, does not induce expression of any of the glucose transporters in the ancestral strains (Figure 3-7B), as expected (Ozcan and Johnston 1999).

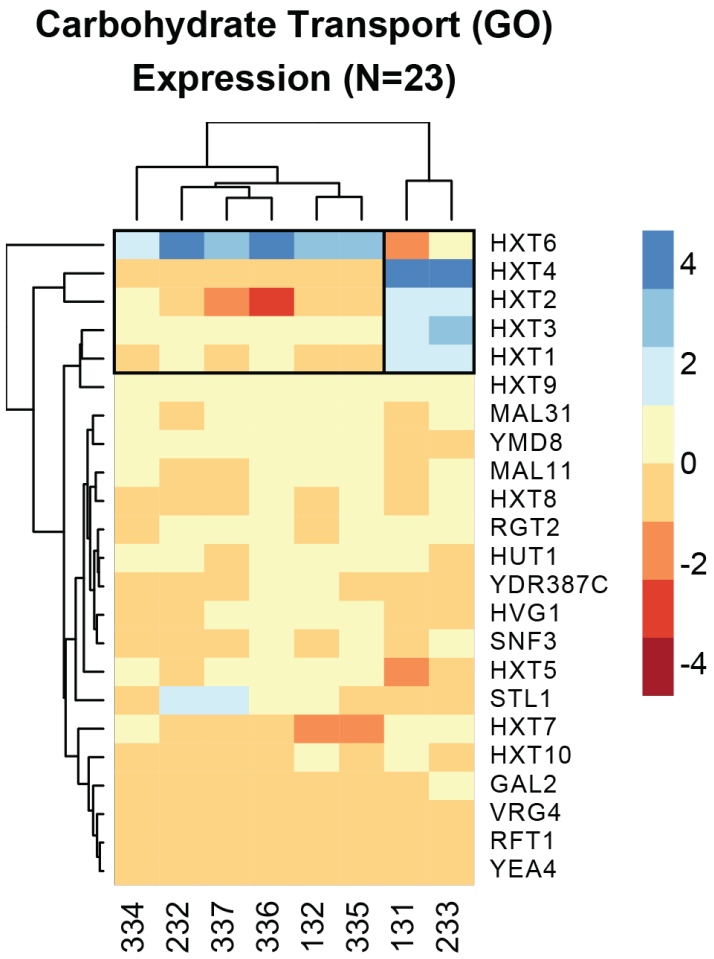


Figure 3-6 Evolved clones differentially regulate carbohydrate transport. Log_2-transformed gene expression of genes annotated with the GO-function of carbohydrate transport. The evolved clones cluster in 2 clades driven by overexpression of either the hexose transporters 1 through 4 (*HXT1-4*) or the *HXT6* transporters. Figure and legend from (Scott, AL et al. 2017).

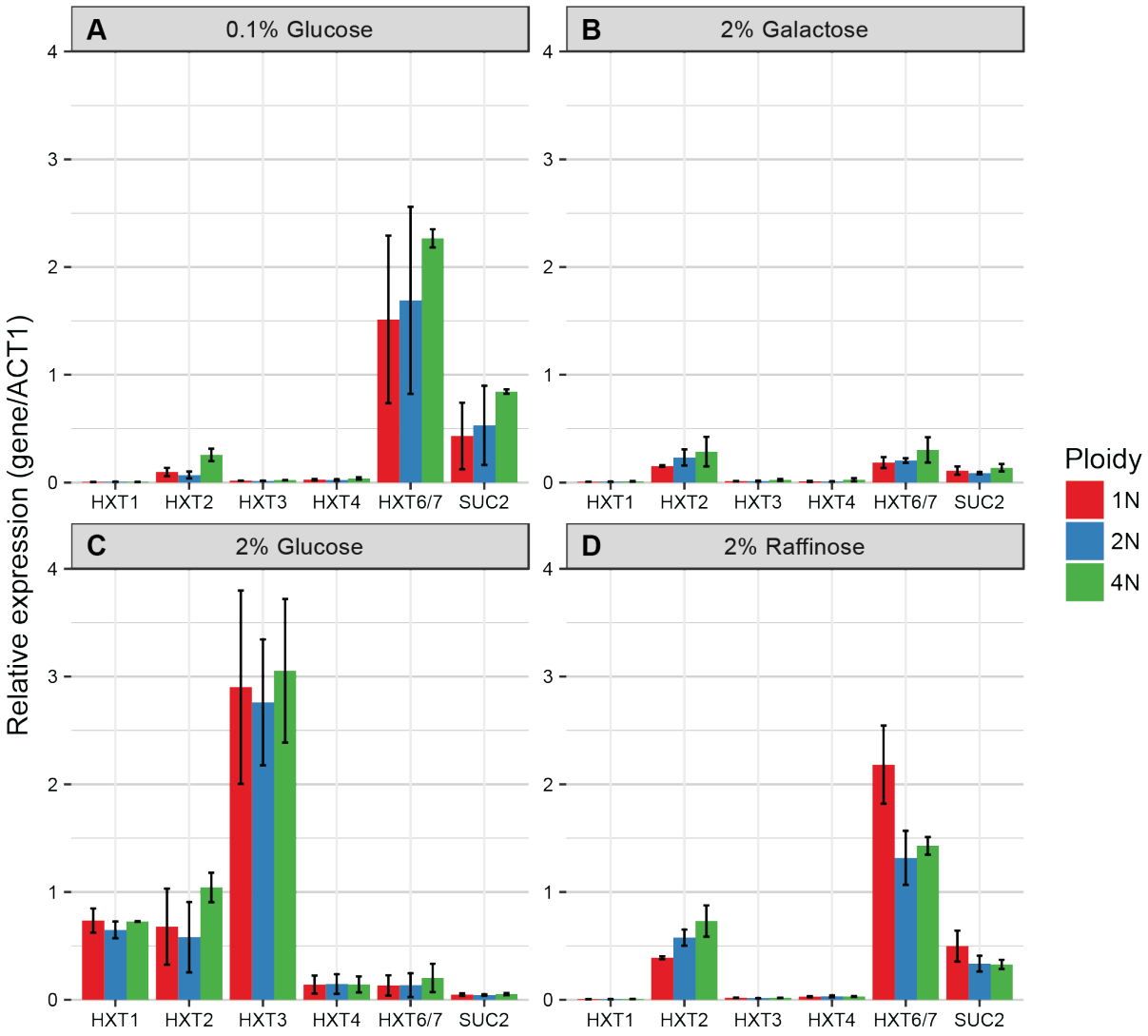


Figure 3-7 Ancestral ploidy does not alter the expression of glucose responsive genes in various carbon sources. qRT-PCR gene expression analysis of glucose responsive genes (*HXT1, HXT2, HXT3, HXT4, HXT6/7*, and *SUC2* internally normalized to *ACT1* expression for the 1N (red), 2N (blue), and 4N (green) ancestral strains grown in A) 0.1% Glucose, B) 2% Galactose, C) 2% Glucose, and D) 2% Raffinose. n=2, error bars represent SEM. Figure and legend from (Scott, AL et al. 2017).

DNA sequence data indicates the strains overexpressing *HXT2*, *HXT3,* and *HXT4* have mutations in glucose sensing and signal transduction (*SNF3* and *MTH1*) while the strains up-regulating *HXT6/7* have amplifications of the *HXT6/7* region (Table 9-1). These data suggest that there are two distinct pathways to up-regulate glucose transport in yeast. This pattern of mutually exclusive transporter usage was observed previously in a study on mutations arising in response to growth in low glucose (Kvitek and Sherlock 2011). The study demonstrated that *MTH1* and *HXT6/7*^{amp} mutations individually give increased fitness, yet a strain with both *MTH1* and *HXT6/7*^{amp} mutations has decreased fitness when grown in a low-glucose environment (defined as negative epistasis).

One exception in the *HXT6/7* expressing cluster is that 4Ne clone 334 does not have *HXT6/7*^{amp}. This strain contains 3 mutations, all of which are located in coding regions and 2 of which are non-synonymous: *YTA7,* a histone binding protein, and *TOR2,* a subunit of the TOR complex (Table 9-1). It is not currently understood how *HXT6* and *HXT7* are fully regulated, therefore it is possible that one of these mutations (*YTA7* or *TOR2*) increases *HXT6/7* expression and may shed light on glucose transport regulation in yeast. Additionally, the TORC/Sch9 pathway has been implicated in genome stability in tetraploid cells and improves fitness in other experimental evolution studies (Lu et al. 2016; Venkataram et al. 2016).

Overall, the RNA expression profiles of the evolved clones indicated that the main mechanism of improved growth in raffinose is to take up more carbon into the cells via hexose transporters. Additionally, there are clearly two mutually exclusive mechanisms to increase hexose transport, by up regulating *HXT2-4* or *HXT6*/*7.* Each of

these pathways can be accessed by mutations in different genes, known to be involved in glucose signaling, such as *SNF3*, *MTH1*, and *HXT6/7*, as well as novel regulators of glucose transport. These data suggest that profiling the expression of a handful of genes may be sufficient to determine the pathway of adaptation in a larger population of the evolved clones.

3.2.4 Evolved clones form distinct clusters by RT-qPCR

To determine whether a small panel of genes can identify the adaptive pathway in an evolved clone, we measured the gene expression pattern of the hexose transporters *HXT2, HXT3, HXT4, HXT6/7*, and the invertase *SUC2* (Table 3-1). This set of genes was, by RNA-seq, diagnostic for the adaptive pathway in a set of evolved clones. We profiled the gene expression panel in 27 evolved haploid, 32 evolved diploid, and 37 evolved tetraploid clones that showed increased fitness in raffinose. RNA was collected from each clone in log-phase growth in raffinose media and gene expression was measured by RT-qPCR.

Hierarchical cluster analysis (HCA) was performed on $\log_2$-transformed normalized gene expression relative to the diploid ancestor grown in the same condition. The resultant gene panel dendrogram is plotted with a heatmap of the gene expression data in Figure 3-8. While the RNA-seq data suggested two distinct patterns of HXT expression, the larger set of evolved clones cluster into five primary clades, labeled *a*-*e*. The optimal number of clades was selected by qualitative inspection of within-group variance estimation versus model complexity (Figure 3-9).

Gene name	Systematic name	Description
HXT2	YMR011W	High affinity glucose and fructose transporter, expression is induced by low levels of glucose and repressed by high levels of glucose.
HXT3	YDR345C	Low affinity glucose and fructose transporter, expression is induced in both low or high glucose conditions.
HXT4	YHR092C	Intermediate affinity glucose and fructose transporter, expression is induced by low levels of glucose and repressed by high levels of glucose.
HXT6/HXT7	YDR343C/ YDR342C	High-affinity glucose and fructose transporters, expressed at high basal levels relative to other HXTs. HXT6 and HXT7 are located adjacent to on chromosome IV and differ by only 2 amino acids.
SUC2	YIL162W	Invertase, the glycosylated form is secreted and hydrolyzes extracellular raffinose into fructose and melibiose.

Table 3-1 Glucose responsive gene expression panel

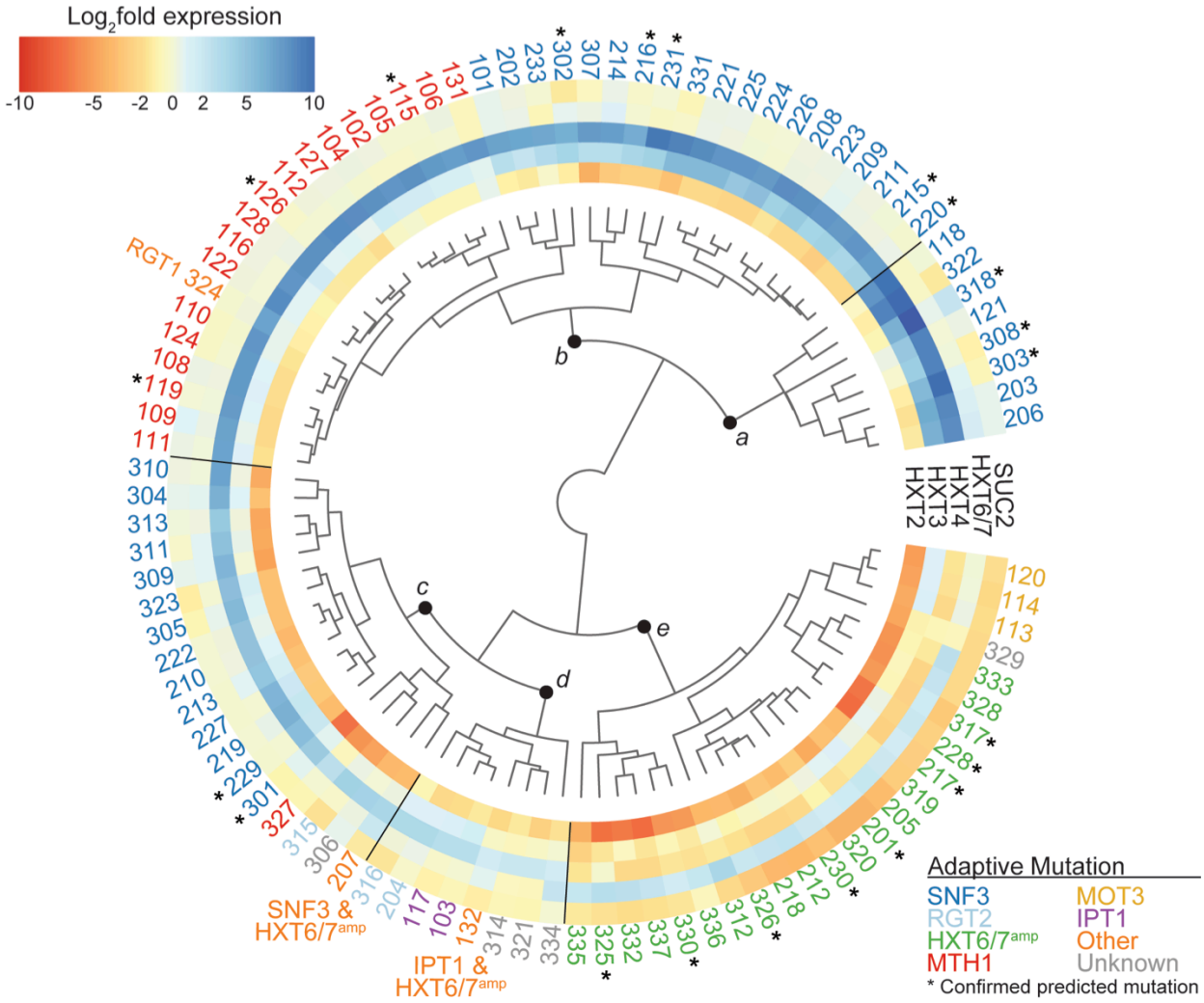


Figure 3-8 Expression of carbohydrate metabolism genes cluster evolved clones by adaptive mutation. Heatmap and dendrogram of log2-transformed qRT-PCR gene expression ratio between haploid evolved clones (n=27), diploid evolved clones (n=32), and tetraploid evolved clones (n=37) and the diploid ancestral strain. Evolved clones are clustered by the expression of a subset of genes involved in carbohydrate transport (*HXT2*, *HXT3*, *HXT4*, and *HXT6/7*) and metabolism (*SUC2*). The strain identifier color indicates the adaptive mutation determined for each evolved clone with whole genome or targeted Sanger sequencing: *SNF3* (blue), *RGT2* (light blue), *MTH1* (red), *HXT6/7*^{amp} (green), *MOT3* (yellow), *IPT1* (purple), other (orange), and undetermined (grey). Asterisks (*) indicate the evolved clones for which targeted Sanger sequencing confirmed the mutation predicted by cluster analysis. Initial ploidy for each evolved clone is indicated by strain number: 1XX =1N, 2XX = 2N, and 3XX = 4N. Figure and legend from (Scott, AL et al. 2017).

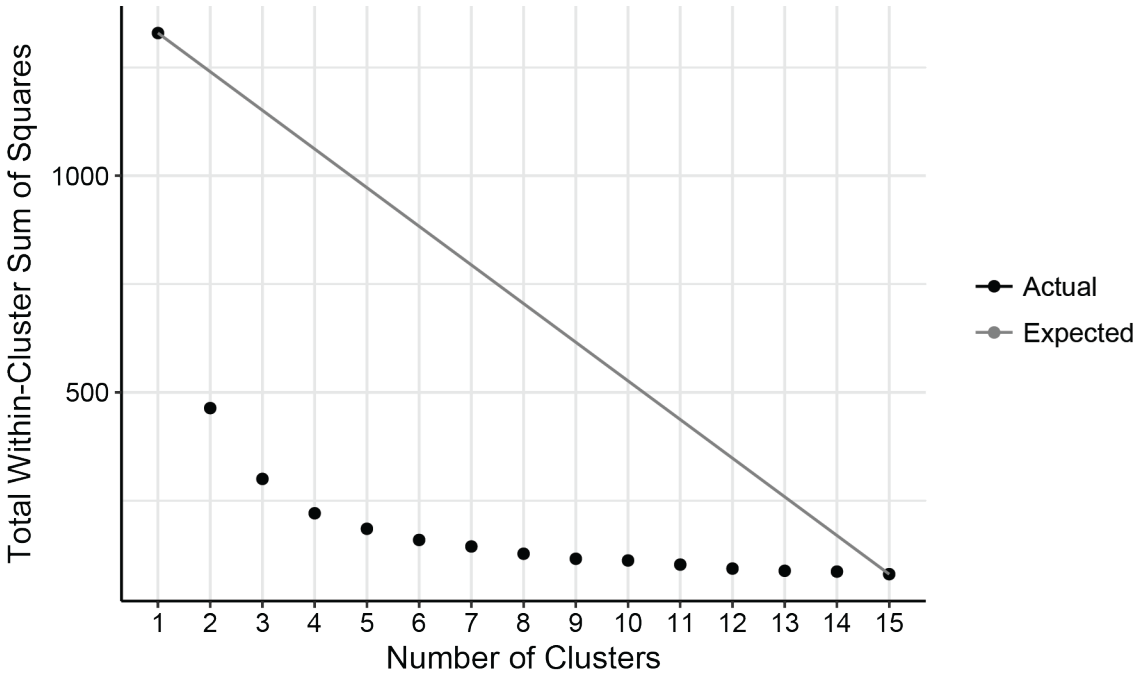


Figure 3-9 Within group variance is minimized at k=5. The averages total sum of squares (TSS) (Y-axis) for k clusters (x axis) was quantified using kmeans for the log_2-transformed gene panel expression data. The TSS was averaged over 20 trials (black dots) at each k. The null hypothesis (grey line) is that the variance decreases linearly with model complexity. The optimal number of groups was determined by visual inspection of the number of clusters (k) at which the change in TSS becomes linear. Figure and legend from (Scott, AL et al. 2017).

3.2.5 Evolved clones form gene expression clusters by adaptive mutation

To understand how the clusters of gene expression related to the underlying mutations in the evolved clones, we determined the most likely primary adaptive mutation for each evolved clone from the WGS data (Table 9-1). Adaptive mutations in genes not previously known to be involved in glucose uptake or metabolism were determined based on multiple strains harboring unique mutations within the same gene. In five evolved clones we were not able to determine a primary adaptive mutation from the WGS. The identity of adaptive mutations in indicated on Figure 3-8 by the color of strain identifier: *SNF3* (blue), *RGT2* (light blue), *HXT6/7*^{amp} (green), *MTH1* (red), *MOT3* (yellow), and *IPT1* (purple), Other (orange) and Unknown (grey). The clades identified by HCA are comprised primarily of evolved clones harboring adaptive mutations within the same gene, despite additional background mutations and underlying karyotype. This is significant because it directly connects the gene expression phenotype in the evolved clones to mutations gained during the evolution experiment.

3.2.6 Gene expression clusters are predictive of adaptive mutation

To test whether our gene expression profiles were predictive of the underlying mutation, we profiled the gene expression panel (Table 3-1) in an additional 3 evolved haploid, 9 evolved diploid, and 9 evolved tetraploid clones for which WGS is unavailable. The adaptive mutation was predicted in each evolved clone based on how the strains cluster in the gene panel dendrogram (Figure 3-8, indicated by an asterisk). The predicted genes were then verified by targeted Sanger sequencing and *HXT6/7* copy number was quantified for all additional strains by either quantitative PCR or aCGH of the genomic DNA. In all twenty-one cases, the predicted mutation was

confirmed (Table 3-2). This represents the first study in which the identity of adaptive mutations was determined based on a particular phenotype, in this case gene expression, in an evolved strain.

Strain Identifier	Gene	Amino Acid	HXT6/7 status
115*	MTH1	I353fs	qPCR No HXT
119*	MTH1	I353fs	qPCR No HXT
126*	MTH1	S133fs	qPCR No HXT
201*	HXT6/7amp		CGH +HXT
215*	SNF3	A214E	qPCR No HXT
216*	SNF3	G157A	qPCR No HXT
217*	HXT6/7amp		CGH +HXT
220*	SNF3	G157A	qPCR No HXT
228*	HXT6/7amp		CGH +HXT
229*	SNF3	E439R	qPCR No HXT
230*	HXT6/7amp		CGH +HXT
231*	SNF3	V470F	qPCR No HXT
301*	SNF3	T385R	CGHed No HXT
302*	SNF3	E550K	CGHed No HXT
303*	SNF3	A491D	qPCR No HXT
308*	SNF3	A491D	qPCR No HXT
317*	HXT6/7amp		CGHed +HXT
318*	SNF3	E413K	CGHed No HXT
325*	HXT6/7amp		qPCR +HXT
326*	HXT6/7amp		qPCR +HXT
330*	HXT6/7amp		CGHed +HXT

Table 3-2 Identity of predicted adaptive mutations

3.2.7 The spectrum of adaptive mutations differs by ploidy level

The primary mode of adaptation at all ploidy levels is to increase glucose uptake by over-expression of the hexose transporters. This is achieved by the up-regulation of *HXT2*, *HXT3*, and *HXT4* primarily through mutations in *SNF3* and *MTH1* or by amplification of the *HXT6* and *HXT7* genomic region. The spectrum of primary adaptive mutations differed in the evolved clones with increasing ploidy level (Figure 3-10). The majority of primary adaptive mutations gained in the haploid strains were LOF mutations in the signal transducer *MTH1*, loss of which is recessive in higher ploidy backgrounds. Conversely, the diploid and tetraploid evolved clones adapted through mutations in glucose sensors or amplification of the glucose transporters *HXT6* and *HXT7.* Overall, the tetraploid evolved clones gained adaptive mutations encompassing more of these genes than haploids and diploids, indicative of greater flexibility in adaptation to raffinose (Figure 3-10).

A significant difference between haploid, diploid, and polyploid yeast is the surface area to volume ratio (Storchová et al. 2006). Since higher ploidy cells have a lower surface to volume ratio, they exhibit differential expression of cell surface constituents, including nutrient transporters (de Godoy et al. 2008; Wu et al. 2010). A decreased surface area to volume ratio at higher ploidy levels is likely to contribute to greater selection acting on the over-expression of these genes (de Godoy et al. 2008). Thus, fundamental differences in cell size with increasing ploidy are one force that may drive the differential spectrum of adaptive mutations with increasing ploidy.

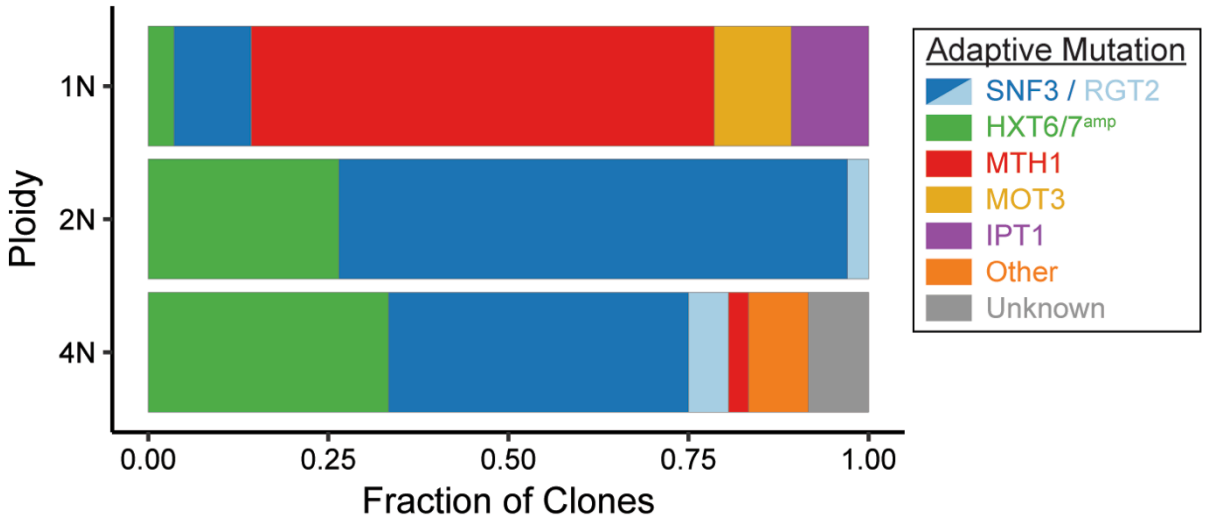


Figure 3-10 The spectrum of adaptive mutations in the evolved clones differs with increasing ploidy. The fraction of evolved clones derived from 1N, 2N, or 4N populations with the indicated adaptive mutation. The color denotes the adaptive mutation determined for each evolved clone with whole genome or targeted Sanger sequencing: *SNF3* (blue), *RGT2* (light blue), *MTH1* (red), *HXT6/7*^{amp} (green), *MOT3* (yellow), *IPT1* (purple), other (orange), and undetermined (grey). Figure and legend from (Scott, AL et al. 2017).

3.2.8 Adaptive mutation is a better separator of evolved clones than ploidy

To examine the effect of underlying ploidy on the gene expression profiles in the evolved clones, we performed linear discriminant analysis (LDA) on the gene expression panel. LDA is a form of dimensionality reduction that is similar to principal component analysis (PCA), which models the data to emphasize the components that explain the most variation in the data set. However, unlike PCA, LDA attempts to model the difference between the data classes as well. We classified the strains by either primary adaptive mutation or initial ploidy (Figure 3-11).

We found that LDA performs considerably better when the data is classified by adaptive mutation rather than initial ploidy. When the strains were classified by adaptive mutation the strains form discrete groups (Figure 3-11). When the evolved clones are classified instead by initial ploidy, the resulting groups were not well defined (Figure 3-11). We did observe some separation of the haploid clones, likely due to the fact that the majority of haploid strains gained a mutation in the same gene, *MTH1*. Together, these results confirm that adaptive mutation is the main driver for expression differences in glucose uptake in the evolved clones. However, the initial ploidy influenced the relative frequency of the different adaptive mutations gained during adaptation.

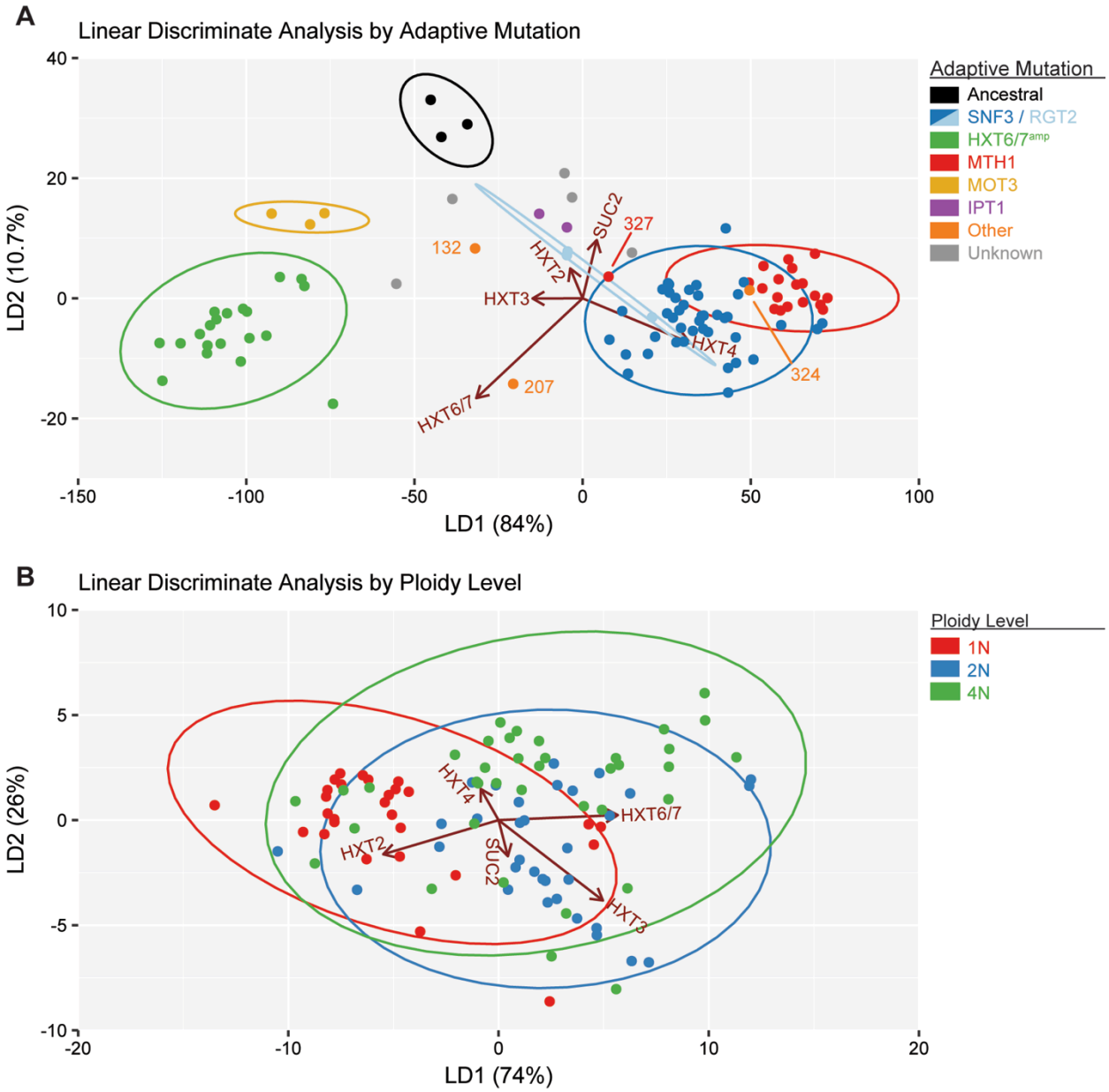


Figure 3-11 Linear discriminant analysis separates evolved clones into distinct populations by adaptive mutation. Linear discriminant analysis (LDA) of the log2-transformed gene expression ratio of the evolved clones relative to the diploid ancestral strain. Gene expression of *HXT2, HXT3, HXT4, HXT6/7*, and *SUC2* was measured by qRT-PCR and normalized to *ACT1* expression. A) LDA of the evolved clones classified by adaptive mutation. The marker color indicates the major classes of adaptive mutations: parental (black), *SNF3* (blue), *RGT2* (light blue), *MTH1* (red), *HXT6/7*^{amp} (green), *MOT3* (yellow), *IPT1* (purple), other (orange), and undetermined (grey). The ellipses denote the confidence interval of 95% for each major class of primary adaptive mutation. B) LDA of the evolved clones classified by initial ploidy. The marker color indicates the initial ploidy: 1N (red), 2N (blue), 4N (green). The ellipses denote the confidence interval of 95% for each ploidy level. Figure and legend from (Scott, AL et al. 2017).

3.3 Materials and Methods

3.3.1 Growth Conditions

Cells were grown for RNA isolations by either Amber Scott or Phillip Richmond. For the RNA expression analyses, yeast were grown to saturation at 30°C at 300rpm for 20 hours in synthetic complete media (SC)+2% raffinose. Cell density was determined by optical density (OD) and cells diluted to 0.2 ODU in SC+2% raffinose. Cultures were grown at 30°C at 300rpm until mid-log growth (0.8-1.0 ODU). RNA-seq cultures were grown in 100mL of media in a flask while qRT-PCR cultures were grown in 5mL culture tubes. Cells were pelleted at 4°C for 5 minutes at 5000xg, washed with DEPC treated water, snap frozen and stored at -80°C.

3.3.2 RNA Isolation

RNA isolations were prepared by Amber Scott and Phillip Richmond. RNA was isolated using the hot acid phenol RNA extraction method and RNA was DNase treated with RQ1 DNase at 37°C for 60 minutes (Promega, M6101). The DNase-treated RNA was extracted with phenol:chloroform:isoamyl alcohol 25:24:1 (sigma, P2069) followed by chloroform extraction. The DNase-treated RNA was precipitated and the pellet was resuspended in 50uL DEPC treated H_20. RNA quality was assed by electrophoresis in MOPs gel or BioAnalyzer (BioRad). Total RNA was used for input into the RNA-seq library and RT-qPCR.

3.3.3 RNA Library Construction

RNA sequencing libraries were prepared by Amber Scott. Strand specific RNA libraries were prepared using the Illumina RNA ligation library protocol from (Levin et al. 2010). Resultant libraries were assessed by BioAnalyzer before sequencing.

mRNA enrichment Poly(A) tailed RNA was isolated from total RNA with the Oligotex mRNA mini kit (Qiagen, 70022) following manufacturer's directions. Unfortunately this method does not sufficiently remove ribosomal RNA. We have since isolated mRNA with magnetic beads, which works much more efficiently. I tested mRNA isolation methods, including the optimal number of bead bindings for the magnetic beads. Details can be found in my lab notebook.

RNA decapping and fragmentation 100ng Poly(A)$^+$ RNA was decapped with 10U tobacco acid pyro-phosphatase (Epicentre, discontinued) with 40U RNAseOut (Invitrogen) for 90 minutes at 37°C in 10ul reaction volume. Decapped-RNA was brought up to a volume of 200ul in DEPC water and cleaned-up with a single phenol:chloroform:isoamyl alcohol extraction (PCIA, 25:25:1, Sigma) followed by chloroform extraction. RNA was precipitated with ethanol for 30 minutes on ice, pelleted and washed with 70% ethanol and resuspended in 16ul DEPC water. Decapped-RNA was fragmented for 6 minutes in 1x fragmentation buffer and fragmented RNA was ethanol precipitated as above and resuspended in 16ul DEPC water.

Dephosphorylation of RNA fragments 3' ends of fragmented RNA was dephosphorylated with 5U Antarctic phosphatase (NEB) with 40U RNAseOut in 1x phosphatase buffer for 30 minutes at 37°C in a 20ul reaction volume and the reaction stopped at 65°C for 5min. 5' ends of fragmented RNA was phosphorylated by adding

20U of T4 polynucleotide kinase (NEB), PNK buffer, 40U RNAseOut, and 5ul 10nM ATP to 3' dephosphorylated RNA for a total reaction volume of 50ul. RNA was phosphorylated at 37°C for 60 minutes. RNA was brought up to 100ul with DEPC water and RNA cleaned up with Rneasy MinElute Qiagen). 350ul RLT butter and 400ul 100% ethanol was added to RNA, mixing by pipetting after each addition, and added to MinElute column, discarding flow through. Column was washed with 500ul RPE. Column dried in fresh 2mL collection tube by spinning at full-speed for 5min with cap open. RNA was collected in 14ul DEPC water in an RNAse-free microfuge tube. Concentrated RNA to a total volume of 6ul in vacufuge.

RNA adaptor ligation The phosphorylated RNA and the 3' sRNA adaptor was denatured at 70°C for 2 min and then chilled the RNA on ice for 2 min. 300U T4 RNA ligase 2, truncated (NEB) with 20U RNAseOut, 0.8ul 100mM MgCl2, and T4 RNA ligase 2 truncated reaction buffer was added to the RNA mix and incubated for 1 hour at 22°C. The 5' was denatured at 70°C before being added to the 3' ligated RNA with 1ul 10mM ATP, 1ul and T4 RNA ligase (NEB) and incubated at 20°C for 1 hour.

First strand cDNA synthesis 3ul of the SRA reverse transcription primer was added to adaptor ligated RNA and RNA denatured at 70°C for 2 min. RNA was reverse transcribed with 600U SuperScript III (Invitrogen) with 30U SUPERase-In (Ambion), 6ul 5x first-strand buffer, and 6ul 100mM DTT for 1 hour at 55°C. The RNA was then degraded with RNAse H for 1 hour at 37°C and RNAse deactivated at 75°C for 15 minutes.

Size selection The remaining cDNA was size selected on a denaturing poly-acrylamide gel for 200-350bp fragments. The gel fragment containing the size-selected

cDNA was crushed and incubated in 500ul 300mM NaCl for 4 hours to overnight at room temperature. The liquid was separated from the gel in a 0.45um cellulose acetate column and cDNA precipitated with ethanol and resuspended in 10ul DEPC water.

PCR amplification The cDNA libraries were amplified with Phusion High-Fidelity DNA polymerase (NEB) and indexing primers compatible with Illumina sequencing. After amplification, libraries were size selected on a non-denaturing polyacrylamide gel for 250-400bp fragments.

3.3.4 RNA Sequencing Analysis

The analysis of the RNA-sequencing libraries was performed by Amber Scott. RNA-seq libraries were sequenced on 1x50 flow cell on an Illumina HiSeq2000 (University of Colorado). Libraries were sequenced to an average depth of 200M reads per strain. Adaptor sequences and low quality reads were trimmed with trimmomatic (v0.32 ILLUMINACLIP:TruSeq2-SE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) (Bolger et al. 2014). There was a considerable amount of ribosomal reads remaining, in order to remove ribosomal reads we first mapped the reads to a custom fasta file containing only the ribosomal gene region (chrXII:450486-459797) with Bowtie2 (v2.02, --sensitive-local) and all unmapped (non-ribosomal reads) were output to an unmapped fastq file (Langmead and Salzberg 2012). The resulting non-ribosomal reads were mapped to *S. cerevisiae* genome (R63.1.1, 2010-01-05) with Bowtie2 (v2.02, --sensitive-local) with the best mapping locations reported for each read (Engel et al. 2014). An average of 58M reads mapped (a 95% mapping rate on non-ribosomal) per strain. SAM files were converted to BAM with samtools v0.1.19 for downstream use (Li et al. 2009). Reads that mapped

unambiguously to transcripts were counted using HTSeq (v0.5.4p5, htseq-count -f bam -s yes -m intersection-strict -t gene -i ID) and the S288c R63-1-1 annotation file; an average of 44M reads per strain mapped to annotated genes.

3.3.5 Differential Expression Analysis

Differential expression analysis was performed by Amber Scott using DESeq v1.10.1 (Anders and Huber 2010). Since there were no biological replicates, the "blind" method was used to estimate dispersions. This uses all samples (strains) to determine the typical variance in expression for each gene. Given the sequenced strains are derived from the same parental strain and contain only a few mutations, we believe this is an acceptable method to estimate dispersions. Differential expression was determined for all evolved clones against the diploid ancestral strain. Genes were considered differentially expressed between two strains if their adjusted p-value was less than 0.05. Complete DESeq output files are included in GEO Accession #GSE95069.

3.3.6 qRT-PCR analysis

Amber Scott performed all qRT-PCR analysis. RNA was reverse transcribed using Multiscribe reverse transcriptase (Thermo Fisher #4311235) with random hexamers. cDNA was diluted to 1:100 and quantified using targeted qPCR primers (Supplementary Materials online) and SYBR select (Life Technologies #4472908) on the Biorad CFX qPCR system. A standard curve was used to determine linear range and efficiency of the primers. Gene expression was internally normalized to *ACT1* expression and error propagated for replicates. Unless otherwise noted, gene expression was measure in triplicate reactions for each of two biological duplicates. To

the best of our ability we attempted to process each or 100 evolved clones and ancestral strains at the same time to prevent variation within a biological replicate.

3.3.7 Cluster analysis

All clustering analysis was performed by Amber Scott on the log_2-transformed qRT-PCR gene expression ratio between the evolved clones and the diploid ancestral strains internally normalized to *ACT1* expression. Hierarchical clustering analysis was performed using the R (v3.3.2, 2016-10-31) hclust function with the "complete" clustering method. Linear discriminate analysis was performed using the lda function in the MASS package (v7.3-45). LDA was performed twice, classifying the strains by either adaptive mutation or by ancestral ploidy (1N, 2N, or 4N).

3.4 Conclusions

In this chapter I determined the key adaptive pathways in the majority of the evolved clones using a panel of 5 genes that was determined to be diagnostic of the pathways of adaptation by RNA-sequencing analysis. I integrated the WGS variant calling with the gene panel expression profile to determine the identity of the adaptive mutations in all but 5 strains. Furthermore, I used the gene expression panel to accurately predict the adaptive mutation in 21 evolved clones for which there was not WGS available. This is the first experimental evolution study to integrate whole genome sequencing with expression analysis to determine the adaptive pathways in a large number of clones isolated from independent populations. The adaptive mutations identified encompass a narrow set of genes, however the higher ploidy strains gain a significantly different spectrum of mutations than haploid strains (Figure 3-10).

I observed frequent parallel adaption in replicate populations, both at the RNA expression level and through mutations in small subset of genes. Strains primarily adapt to growth in raffinose media by up-regulating the hexose transporters that are required to import glucose (or fructose) into the cell. The majority of strains either acquired mutations in glucose sensors or signal transducers that led to the over-expression of *HXT2, HXT3,* and *HXT4* or amplified the *HXT6/7* genomic region leading to the over-expression of *HXT6/7*. Interestingly, gene expression patterns for all evolved clones cluster according to one key adaptive mutation that they carry, despite additional background mutations and underlying karyotype or ploidy level (Figure 3-8, Figure 3-11). While I find that ploidy level itself is not a good predictor of the adaptive pathways (Figure 3-11), the tetraploid clones have increased genetic complexity relative to haploids and diploids, suggesting that in this relatively short evolutionary timescale polyploid cells explored more evolutionary innovations.

Strains with mutations in *SNF3* and *MTH1,* while discrete, are grouped closely together by LDA, indicating that they have a similar expression profile (Figure 3-11). Despite the apparent parallel adaptation in these strains at the gene expression level, the nature of the acquired mutations (*MTH1* and *SNF3*) is drastically different. The dominance of a mutation is the key determinant of diploid and tetraploid fitness benefit and evolvability compared to haploid lineages (Otto and Whitton 2000; Gerstein et al. 2011). Thus, differing levels of dominance and differing pleiotropic effects of mutations in *MTH1* or *SNF3* may affect on the long-term evolutionary outcomes in these strains.

Additionally, amplifications of the *HXT6/7* region occurred more often with increasing ploidy level. Unlike single nucleotide polymorphisms or indels, $HXT6/7^{amp}$

copy number can change rapidly (Møller et al. 2015; Mishra and Whetstine 2016). The ability to rapidly respond to environmental changes in glucose concentration represents a significant advantage to strains that gained *HXT6/7*^{amp}*.* Thus, the higher likelihood of acquiring *HXT6/7*^{amp} in the 2Ne and 4Ne evolved populations compared to the 1Ne population represents a significant advantage to higher ploidy during adaptation.

While the majority of adaptive mutations occurred in genes with known roles in glucose uptake, I also recovered adaptive mutations in genes with no previously known to have a role in glucose uptake or metabolism (*MOT3* and *IPT1*). Strains with mutations in these novel genes formed distinct clusters in HCA and LDA of the glucose response gene panel, suggesting that the mechanisms of adaption in these strains have some role in glucose uptake. Additionally, I found dominant mutations in the glucose signal transducers *RGT1* and *MTH1,* previously thought to only acquire recessive mutations in glucose-limited environments. Further characterization of the adaptive mutations (Chapter 4) is required to fully understand the mechanisms of adaptation of yeast of different ploidy levels to growth in raffinose media.

4 CHARACTERIZATION OF ADAPTIVE MUTATIONS

Portions of this chapter are published previously or are currently under review:

1. Anna M. Selmecki, Yosef E. Maruvka, Phillip A. Richmond, Marie Guillet, Noam Shoresh, **Amber L. Sorenson**, Subhajyoti De, Roy Kishony, Franziska Michor, Robin Dowell & David Pellman. 2015. Polyploidy can drive rapid adaptation in yeast. Nature 519:349–352.
2. **Amber L Scott**, Phillip A. Richmond, Robin Dowell, Anna M. Selmecki. The influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source. *Under Review.*

Anna Selmecki constructed the SNF3 mutant strains and chromosome XIII aneuploid strains. She also measured the relative fitness of these strains and the evolved clones, the results of which are summarized in this chapter.

4.1 Introduction

To date, experimental microbial evolution studies have focused on the rate of adaptation, the number of mutations acquired, and population dynamics with very little attention given to the nature of the particular mutations gained. In fact, so little attention is paid to the mutations themselves that studies have gone as far as to incorrectly assume the class (loss-of-function vs. gain-of-function) of specific mutations (Koschwanez et al. 2013; Kvitek and Sherlock 2013). This may be, in part, because the majority of experimental evolution studies have been conducted with haploid organisms. To fully understand the mechanisms of evolution, a deeper understanding of the nature of beneficial mutations is required. In this chapter I characterize the key adaptive

mutations recovered in our study. I give special attention to the impact of ploidy level of the relative frequency and beneficial effect of the mutations.

The key adaptive mutations recovered in our study occur primarily in genes with a known role in glucose sensing and uptake; namely, glucose sensors, signal transducers, and glucose transporters. However, the nature of adaptive mutations in these genes is understudied. For a summary of glucose sensing and uptake in yeast see Chapter 1.8. Here we explore the adaptive nature of a number of mutants recovered from our experimental evolution. I also characterize adaptive mutations recovered in genes with no canonical role in carbon utilization (*IPT1* and *MOT3*). Finally, I examine the effect of chromosome XIII aneuploidy at differing ploidy levels.

4.2 Results and Discussion

4.2.1 Evolved clones exhibit dramatic fitness gains after 250 generations in raffinose media

The ancestral clones exhibited dramatic fitness gains relative to the diploid ancestor after growth in raffinose media for 250 generations. To understand the impact of ploidy level and the specific adaptive mutations on fitness gains, we plotted the competitive fitness for each adaptive mutation in the 1Ne (Figure 4-1A), 2Ne (Figure 4-1B), and 4Ne (Figure 4-1C) evolved clones. Competitive fitness was measured for each strain relative to the diploid ancestral strain. The 1N, 2N, and 4N ancestral strains had an average competitive fitness of 0.33, 0, and -1.68 relative to the diploid ancestor, respectively. While the 4Ne clones have the lowest competitive fitness after 250 generations (Figure 4-1C), they made the largest fitness gains. Thus, tetraploids not only adapt more quickly than haploids and diploids, they also achieve greater fitness

gains. This suggests that polyploidization events may be a critical factor for improving evolutionary outcomes during adaptation to certain growth conditions.

We further asked if the specific adaptive mutations impacted the relative fitness of the evolved clones. The levels of competitive fitness are largely the same among all evolved clones after 250 generations (Figure 4-1). The one exception is 1Ne clones with mutations in *MOT3* have significantly greater relative fitness than the 1Ne clones with mutations in *MTH1*. This is surprising given the relatively unknown role for *MOT3* in adaptation to growth in carbon-limited media. The *MOT3* mutations are examined in greater detail in Chapter 4.2.6. These data suggest that the fitness outcomes are independent of adaptive mutation, at least in raffinose media. However, the long-term benefit of particular mutations will depend on dominance level of the mutation and its pleiotropic (i.e. off-target) effects in changing environments (Adams and Rosenzweig 2014; Bleuven and Landry 2016).

Previous experimental evolutions studies of glucose- limited growth in chemostats (Figure 1-1) found evidence for clonal interference between clones in the population (Kao and Sherlock 2008). Clonal interference is the phenomenon in which independent beneficial mutations arise within large, usually asexual, populations and compete for resources. Competing beneficial mutations are theorized to slow the overall rate of evolution (Kao and Sherlock 2008). While we do not have evidence of clonal interference in our evolved populations, the relatively similar competitive fitness levels in the clones that harbor beneficial mutations in different genes suggest that clonal interference could indeed prevent the expansion of a particular lineage. However, head-

to-head competitions between strains with differing beneficial mutations are necessary to fully understand the contribution of clonal interference to the rate of adaptation.

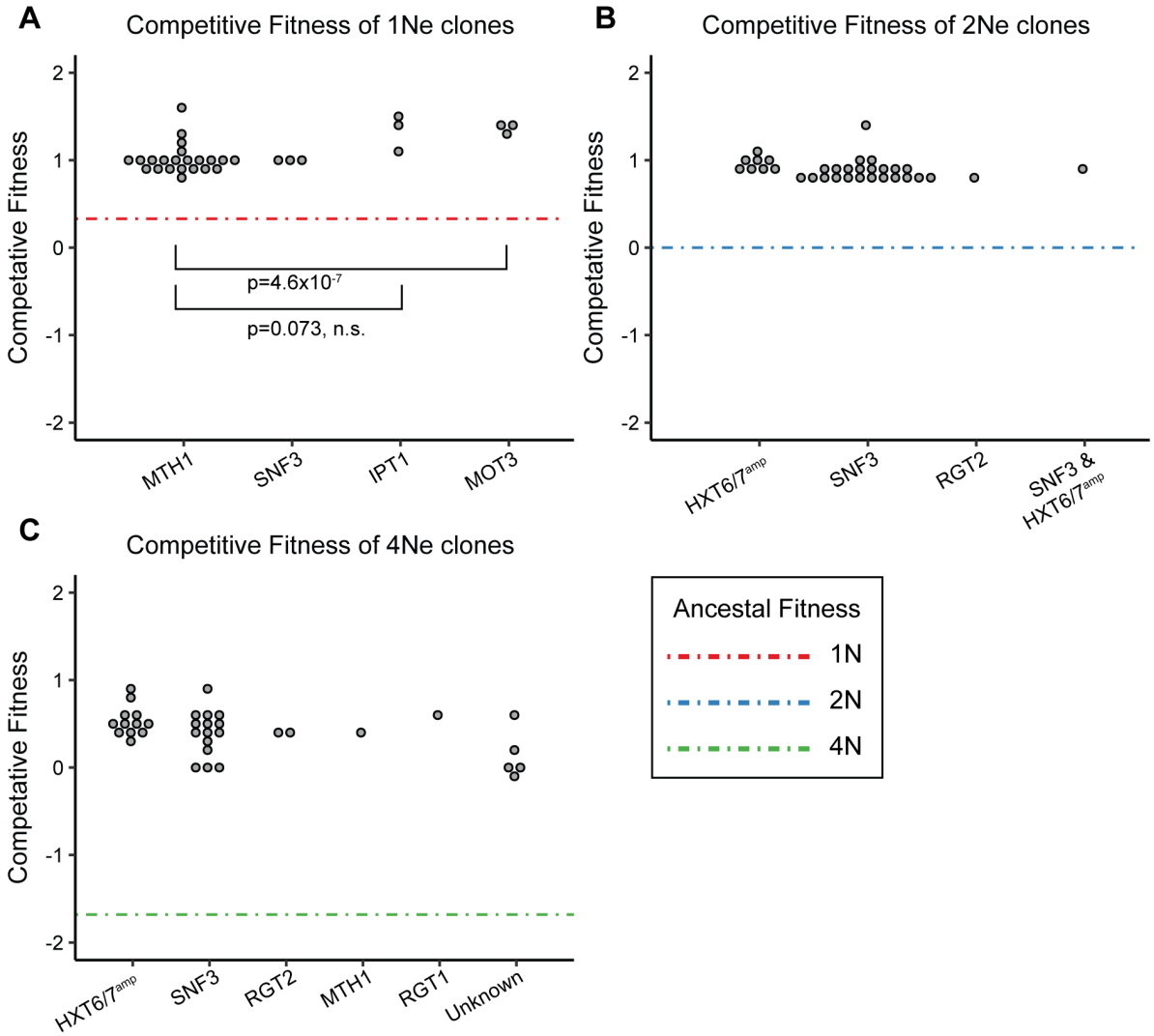


Figure 4-1 Evolved clones exhibit increased fitness in raffinose media. Competitive fitness of the evolved clones isolated from A) 1Ne, B) 2Ne, C) and 4Ne populations. Competitive fitness (Y-axis) is relative to the diploid ancestral strain. The evolved clones are separated by primary adaptive mutation (X-axis). The dashed lines indicated starting fitness of the ancestral 1N (red), 2N (blue), and 4N (green) strains. 1Ne clones with mutations in *MOT3* have significantly greater competitive fitness than the 1Ne clones with mutations in *MTH1*, paired student t-test. Figure and legend from (Scott et. al. 2017).

4.2.2 Activating mutations in the glucose sensors, Snf3p and Rgt2p

Two closely related glucose sensors, Snf3p and Rgt2p, signal through downstream proteins to activate expression of the HXT genes in conditions of low and high extracellular glucose levels, respectively (Figure 1-3)(Sabina and Johnston 2009). *SNF3* was the most commonly mutated gene in the evolved clones. Mutations in *SNF3* were most often found in the 2Ne strains (18 of 24) followed by 4Ne strains (10 of 28) but only rarely found in the 1Ne (3 of 24) (Table 9-1). A small number of strains harbor mutations in *RGT2*: 1 of 24 in 2Ne clones and 2 of 28 in 4Ne clones (Table 9-1). Raffinose acts as a source of low glucose; therefore it makes sense that there is a lower frequency of mutations in the high glucose sensor Rgt2p than in the low glucose sensor Snf3p.

Strains with mutations in these sensors clustered into multiple clades within the gene panel dendrogram (Figure 3-8). The strains with mutations in the gene encoding the glucose sensor Snf3p grouped within clades *a, b, and c* and vary in their levels of expression of *HXT2, HXT3* and *HXT4*. Strains in clade *a* show robust overexpression of *HXT3* and *HXT4,* while strains in *clade b* have moderate overexpression of *HXT3* and high expression of *HXT4.* Clade *c* is characterized by minimal up regulation of *HXT3* and *HXT4*, and down regulation of *HXT2.* The strains with mutations in *RGT2* are primarily found in *clade d* and have only minimal over expression of *HXT3* and *HXT4* when compared to the diploid ancestor. These expression patterns are consistent with the activation of glucose responsive signaling through Snf3p or Rgt2p. Moreover, evolved clones with *SNF3* mutations exhibit glucose independent activation of the hexose transporters (Figure 4-2) (Pasula et al. 2007).

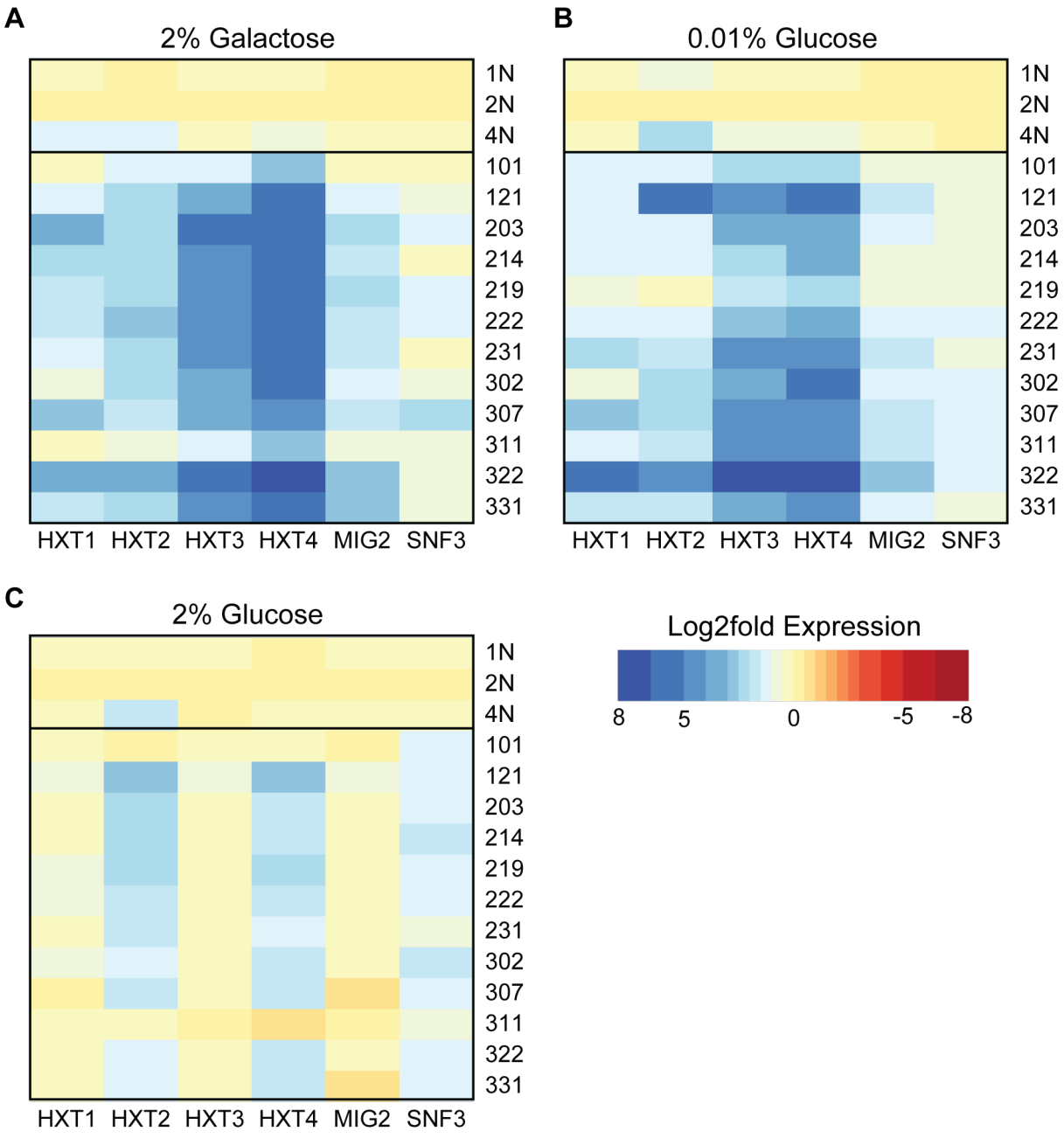


Figure 4-2 Mutations in SNF3 result in constitutive activation of glucose responsive genes. qRT-PCR expression analysis of glucose responsive genes normalized to ACT1 for selected evolved clones with mutations in SNF3 relative to the diploid ancestral strain. Strains were grown 0.1% glucose overnight and then grown in A) 2% Galactose, B) 0.1% Glucose, and C) 2% Glucose for 6 hours or until cultures reached an OD of 0.8. Figure and legend from (Scott et. al. 2017).

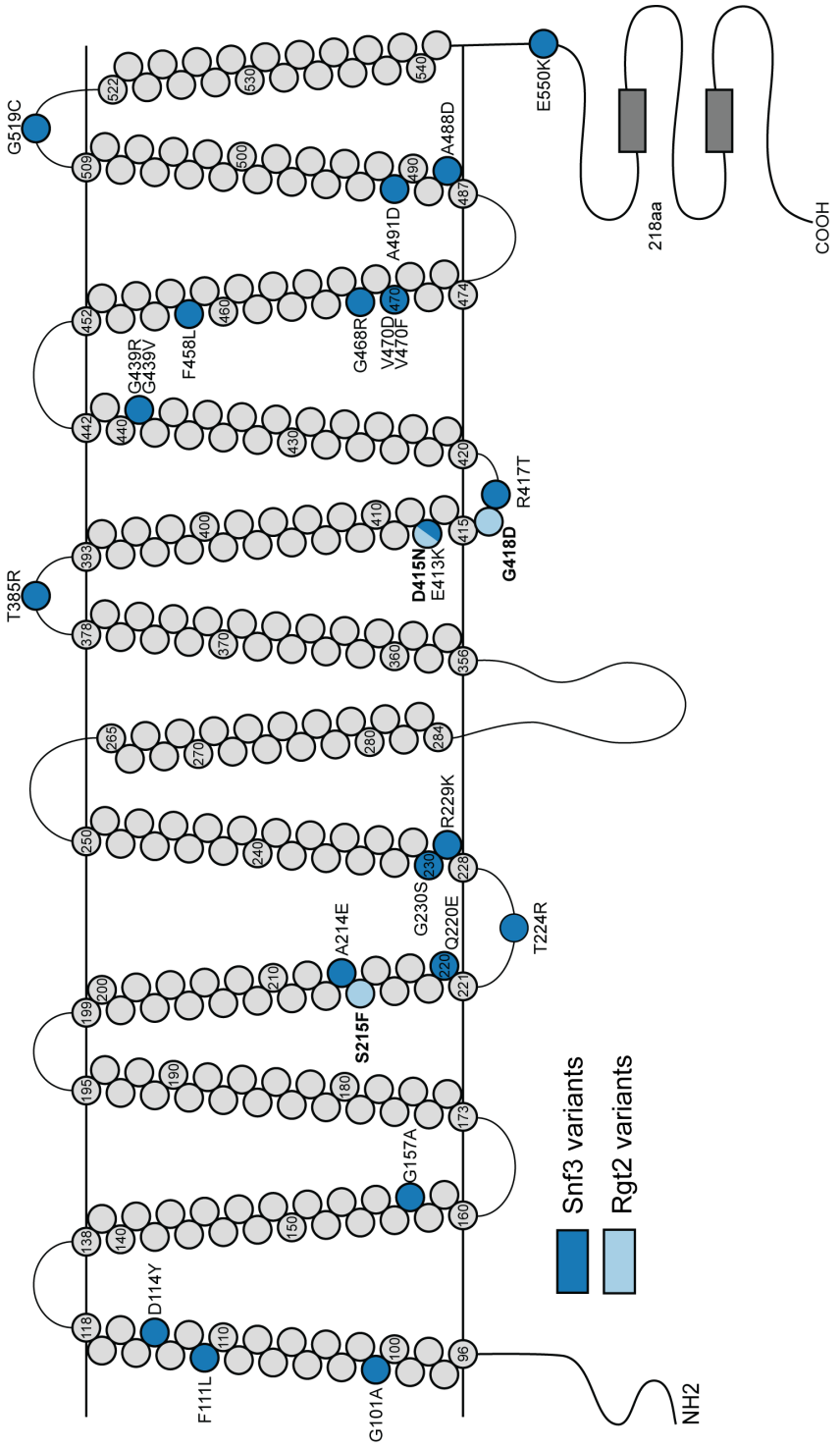


Figure 4-3 Dominant mutations in the glucose sensors, Snf3 and Rgt2, were found in evolved clones derived from each ploidy. Diagram of Snf3 (blue) and Rgt2 (light blue, italic) amino-acid substitutions that resulted from single nucleotide variations recovered in the evolved clones. The locations of transmembrane, cytosolic, and extracellular domains were predicted with TMHMM 2.0c (Krogh et al. 2001). Figure and legend from (Scott et. al. 2017).

The different clusters formed by *SNF3* mutants in the gene panel dendrogram are indicative of varying levels of downstream activation: high (clade *a*), medium (clade *b*), and low (clade *c*). We hypothesized that the level of activation may be a result of the location and identity of the mutation in Snf3p. To test this, we diagramed the locations of the mutations in Snf3p (Figure 4-3). In general, mutations with high activation were located in transmembrane domains 8, 10, and 11, near the cytoplasmic domains. Mutations in strains that cluster with low activation are located in the extracellular and cytoplasmic loops. Other possible factors may contribute to downstream activation, such as other background mutations or underlying karyotype. Further tests are required to elucidate the exact mechanisms of varying levels of Snf3p signal activation.

To test the impact of *SNF3* mutations at different ploidy levels, in the absence of other background mutations, we constructed isogenic *SNF3-G439E* strains differing only by ploidy. We found that *SNF3–G439E* had a dominant, raffinose-specific, beneficial effect that was relatively stronger in the 4N strain (Figure 4-4; t-test, $p<1x10^{-4}$). We measured the expression of the glucose responsive gene panel (Table 3-1) in the *SNF3-G439E* ploidy series. These strains exhibit an expression signature similar to the evolved clones harboring mutations in *SNF3*, confirming that the gene expression pattern is a result of the mutations in *SNF3* (Figure 4-5). Moreover, the extent of HXT overexpression may be mediated by the ratio of mutant alleles to wildtype alleles: the haploid and homozygous diploid *SNF3–G439E* strains have increased *HXT2-4* expression (Figure 4-5) relative to the diploid ancestor. Together, these data suggest that underlying karyotype may modulate the beneficial effect of *SNF3* mutations.

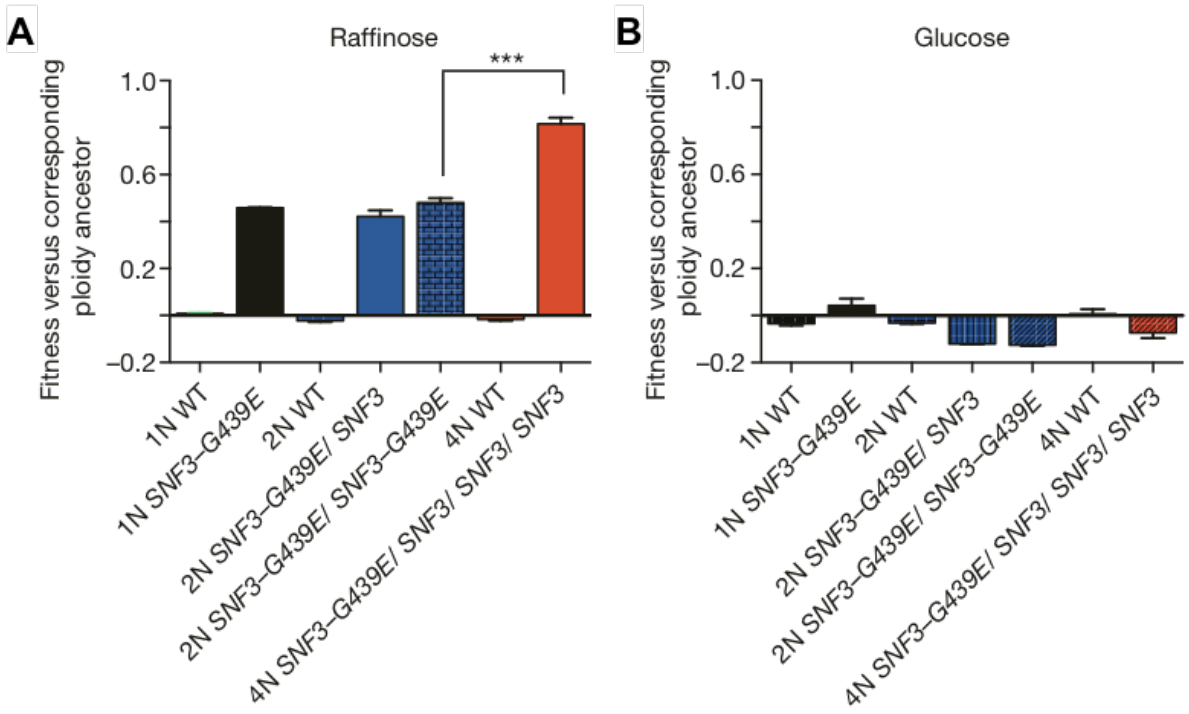


Figure 4-4 Mutations in SNF3 provide a raffinose specific fitness benefit. Competitive fitness of engineered isogenic strains of the indicated ploidy and genotype, relative to the corresponding ploidy ancestor, in A) raffinose and D) glucose medium. Error bars, mean +/- SEM of three independent SNF3–G439E transformants of each ploidy type, t-test ***$p<1x10^{-4}$. Figure and legend from (Selmecki et al. 2015).

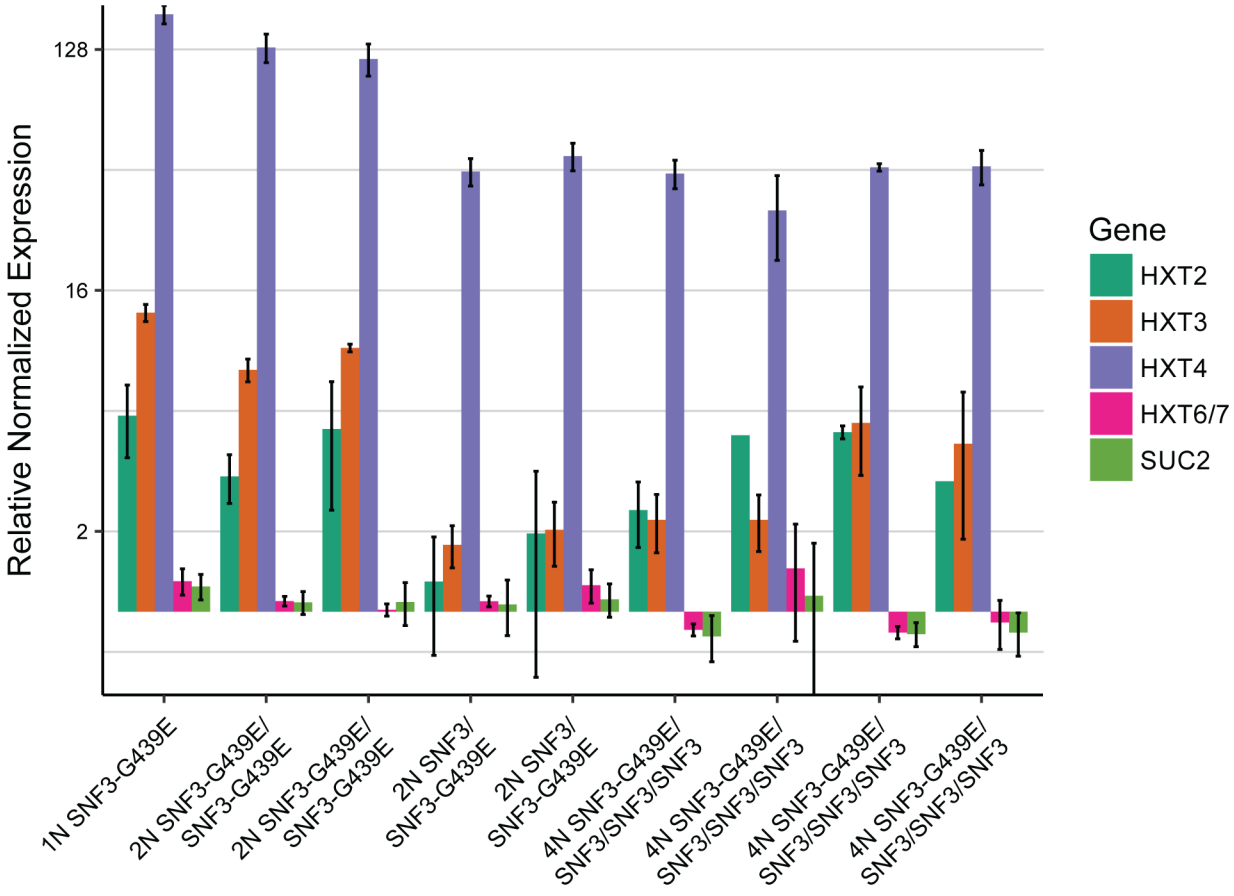


Figure 4-5 SNF3 mutant copy number impacts HXT expression. Expression of the glucose responsive gene panel in engineered isogenic strains of the indicated ploidy and genotype, relative to the diploid ancestor. RT-qPCR expression values are internally normalized to *ACT1*. Y-axis is in Log_2 scale, n=3, error bars represent SEM.

One determinant of long-term evolutionary success is ability to adapt to changing environments. Mutations that may be beneficial to growth in raffinose may show trade-offs in other carbon environments (Wenger et al. 2011; Adams and Rosenzweig 2014; Bleuven and Landry 2016). *SNF3–G439E* mutations show a slight decrease in fitness when grown in rich media (Figure 4-4). Moreover, the mutations recovered in *SNF3* increase downstream signaling to levels above and beyond the levels evolved in the ancestral strain. This suggests that natural selection restricted the signaling capabilities of Snf3p, indicating that increased activation of the glucose responsive genes may be selected against in typical growth conditions or the changing environments in which yeast evolved. A similar effect was discovered on the gene *SFA1,* mutations of which increased resistance to formaldehyde (Zhang et al. 2013).

4.2.3 Recessive and dominant mutations in the signal transducer, Mth1

Under repressing conditions, Mth1p binds Rgt1p at the promoters of glucose inducible genes, repressing transcription though Ssn6p/Tup1p (Kim et al. 2003; Polish et al. 2005; Roy et al. 2013). However, in the presence of low extracellular glucose or fructose concentrations, Snf3p signals the degradation of Mth1p (H Liang and Gaber 1996; Lafuente et al. 2000; Flick et al. 2003). This results in Rgt1p phosphorylation and the de-repression of glucose inducible genes (Polish et al. 2005; Pasula et al. 2010).

Several of the evolved clones have mutations in the signal transducer *MTH1* (Table 9-1). With one exception, these clones all cluster together as a sub-clade within clade *b* (Figure 3-8), consistent with the fact that Snf3p signals downstream for the degradation of Mth1p to relieve the repression of the hexose transporters. These strains all moderately overexpress *HXT3* and highly overexpress *HXT4.* The major

outlier is the 4Ne clone 327, which harbors a dominant mutation in *MTH1* and clusters in clade *c.* This dominant *MTH1* mutation results in a heterozygous C321F substitution (Figure 4-6, Table 4-1). This strain has a different expression profile than the other *MTH1* mutant strains, suggesting that this mutation may not fully abolish *MTH1* function. While, dominant non-degradable *MTH1* mutants that result in constitutive repression of the hexose transporters have been previously described (Schulte et al. 2000), this mutation is the first known dominant activating mutation in *MTH1*.

Mutation	Background	Strain	Experimental details	Citation
H154Tfs	W303 (1N)	EvoClone2	1mM Sucrose, serial passage, POL3-L523D mutant	Koschwanez et al. 2013
S51Ifs	W303 (1N)	EvoClone6	1mM Sucrose, serial passage, POL3-L523D mutant	Koschwanez et al. 2013
W427C	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
Q416K	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
Y332Stop	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
C321W	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
Y311Stop	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
Y245Stop	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
Q236Stop	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
L156Stop	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
S106Stop	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
N314H	S288C (1N)	E2, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
R385Stop	S288C (1N)	E2, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
E249Stop	S288C (1N)	E2, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
C188Stop	S288C (1N)	E2, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
S101Stop	S288C (1N)	E2, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
K86Stop	S288C (1N)	E2, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013

Q374Stop	S288C (1N)	E3, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
Q338Stop	S288C (1N)	E3, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
L241Stop	S288C (1N)	E3, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
S209Stop	S288C (1N)	E3, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
1353fs	S288C (1N)	102	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	104	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
N12fs	S288C (1N)	105	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	106	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	108	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
K333Stop	S288C (1N)	109	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
N69fs	S288C (1N)	110	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
K367fs	S288C (1N)	111	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
S133fs	S288C (1N)	112	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	116	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	122	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	124	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
A475D	S288C (1N)	127	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	128	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	131	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
C321F	S288C (4N)	327	2% Raffinose, serial passage, 250 generations	Selmecki et. al. 2015
1353fs	S288C (1N)	115	2% Raffinose, serial passage, 250 generations	Scott et. al. 2017
1353fs	S288C (1N)	119	2% Raffinose, serial passage, 250 generations	Scott et. al. 2017
S133fs	S288C (1N)	126	2% Raffinose, serial passage, 250 generations	Scott et. al. 2017

Table 4-1 *MTH1* mutations recovered from experimentally evolved yeast populations under carbon stress.

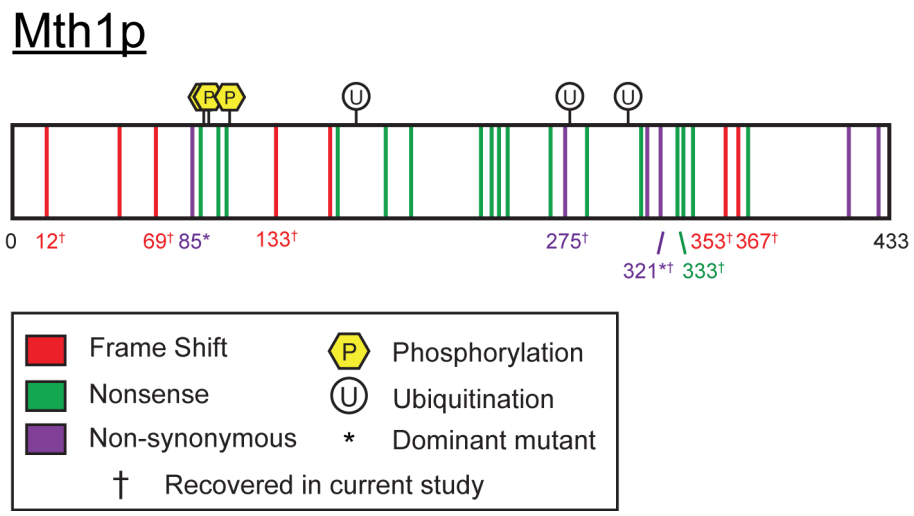


Figure 4-6 Dominant and recessive mutations have been recovered in MTH1 in glucose-limited growth conditions. Diagram of published mutations in MTH1. The type of mutation is indicated by color: frameshift (red), nonsense (green), and non-synonymous (purple). Known ubiquitiation (white, circle) and phosphorylation (yellow, hexagon) are also indicated. Figure and legend adapted from (Scott et. al. 2017).

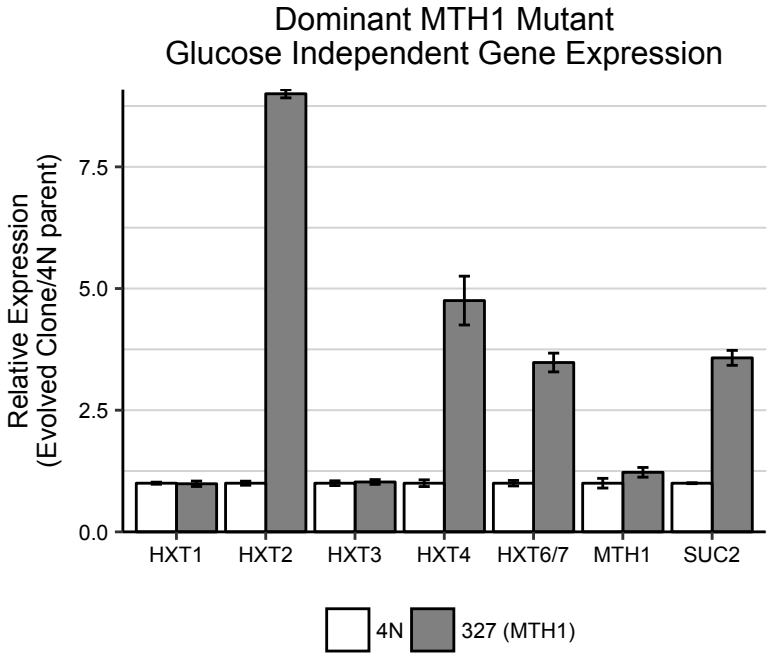


Figure 4-7 MTH1-C321F induces glucose independent expression of glucose responsive genes. qRT-PCR gene expression analysis of glucose responsive genes relative to *ACT1* expression and normalized to the 4N ancestral strain grown in 2% galactose. n=2, error bars represent SEM. Figure and legend adapted from (Scott et. al. 2017).

When grown in media with galactose as the sole source of carbon, 4Ne clone 327 exhibits glucose independent activation of genes that are typically repressed by Rgt1p in glucose-free conditions (Figure 4-7) (Ozcan and Johnston 1995). This suggests that a single copy of mutant Mth1p is acting ether as a dominant negative, by preventing wildtype Mth1p from binding Rgt1p, or more likely preventing proper function of Rgt1p. Mth1p^{C321F} may prevent proper Rgt1p function in two ways: by preventing Rgt1p binding at the promoters of the glucose responsive genes or preventing recruitment of the repressive complex Ssn6p/Tup1p to the promoters of glucose responsive genes. It has been reported that recruitment of Ssn6p/Tup1p is required for the release of Rgt1p from binding the *HXT1* promoter and the interaction of Rgt1p and Ssn6p/Tup1p is mediated by Mth1p (Roy et al. 2013; Roy et al. 2014). Thus, I hypothesize that Mth1p^{C321F} disrupts the interaction between Rgt1p and Ssn6p/Tup1p, both preventing repression of the glucose responsive genes and causing non-reversible binding of Rgt1p/Mth1p^{C321F} at the promoters of glucose responsive genes. However, functional studies of the *MTH1-C321F* mutation are required to elucidate the effect of the mutations on glucose responsive gene regulation.

4.2.4 A novel dominant mutation in the transcriptional repressor, Rgt1p

We sequenced a single evolved clone with a mutation in the negative regulator of glucose responsive gene expression, *RGT1* (Table 9-1). This strain, 4Ne clone 324, clusters with the *MTH1* mutants in cluster *b* (Figure 3-8). The clustering of this strain in clade *b* is consistent with loss of Rgt1p function in repression of the glucose responsive genes, yet this mutation occurs in only 1 of 2 alleles of a tetraploid evolved strain. Importantly, it has been previously reported that loss of a single *RGT1* allele does not

result in haploinsufficiency (Dietzel et al. 2012). The dominant mutation that we recovered (*rgt1*^{S509stop}*/RGT1*) leads to the dominant activation *HXT2* and *HXT4* in galactose (Figure 4-10).

The location of the dominant mutation is intriguing, as it results in a heterozygous truncation of the domain of Rgt1p (Figure 4-8, asterisk, Table 4-2) previously described as an inhibitor of repression (Polish et al. 2005). This domain is thought to function as a self-inhibiter of the DNA binding domain of Rgt1p. Thus, we would expect that loss of this region would result in constitutive binding of Rgt1p and, as a result, constitutive repression. However, we see the opposite effect, the expression profile of *rgt1*^{S509stop}/*RGT1* phenocopies loss of *MTH1*. Previous experimental evolution studies on haploid yeast populations in glucose-limiting conditions recover mutations in *RGT1* that are presumed to be LOF (Koschwanez et al. 2013; Kvitek and Sherlock 2013). However, the reported mutations within *RGT1* were located in the same inhibitor of repression domain (Figure 4-8), suggesting they were unlikely LOF.

Since we did not recover haploid LOF mutations in *RGT1*, we hypothesized that loss of Rgt1p would be detrimental to growth in raffinose. To test this, we deleted *RGT1* in the haploid ancestral strain. Multiple *rgt1*Δ clones displayed a growth defect on raffinose compared to the 4Ne clones 324 (Figure 4-9B). We did not observe a difference in growth for *rgt1*Δ strains compared to the haploid ancestor in raffinose or glucose (Figure 4-9). These results suggest that the *rgt1*^{S509stop}/*RGT1* mutation in 4Ne clone 324, as well as mutations in *RGT1* described in other studies, do not fully abolish Rgt1p function. Instead, these mutations may affect association with co-regulators, such as Mth1p.

Mutation	Background	Strain	Expermintal details	Citation
L832fs	W303 (1N)	EvoClone4	1mM Sucrose, serial passage, POL3-L523D mutant	Koschwanez et al. 2013
Q1053Stop	W303 (1N)	EvoClone1	1mM Sucrose, serial passage, POL3-L523D mutant	Koschwanez et al. 2013
G687V	W303 (1N)	EvoClone10	1mM Sucrose, serial passage, POL3-L523D mutant	Koschwanez et al. 2013
Y526C	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock pgen 2013
V565L	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
L575W	S288C (1N)	E1, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
C646F	S288C (1N)	E2, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
E817K	S288C (1N)	E2, Bulk	0.08% Glucose, Chemostat, 448 generations	Kvitek and Sherlock 2013
S509Stop	S288C (4N)	324	2% Raffinose, serial passage, 250 generations	Selmecki et al. Nature 2015

Table 4-2 *RGT1* mutations recovered from experimentally evolved yeast populations under carbon stress.

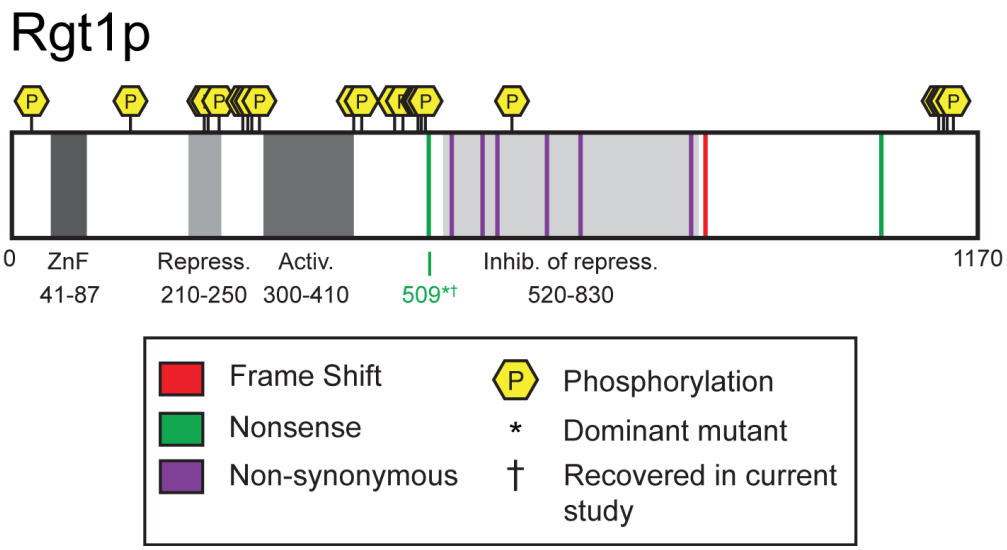


Figure 4-8 Diagram of published mutations in RGT1. The type of mutation is indicated by color: frameshift (red), nonsense (green), and non-synonymous (purple). Known ubiquitiation (white, circle) and phosphorylation (yellow, hexagon) are also indicated. Figure and legend adapted from (Scott et. al. 2017).

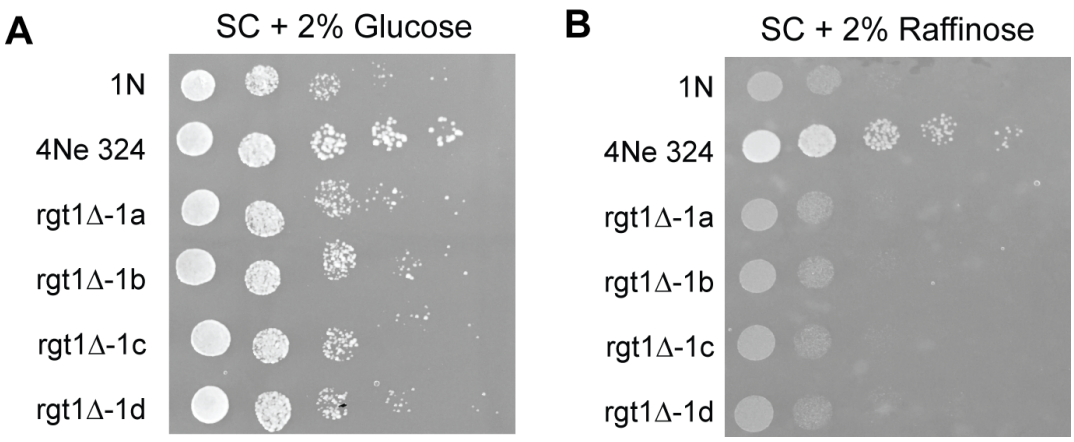


Figure 4-9 r*gt1Δ* mutants exhibit poor growth on raffinose media. Spot assay for *rgt1Δ* strains grown on synthetic complete agar plates containing either A) 2% glucose or B) 2% raffinose as the sole carbon source. Figure and legend adapted from (Scott et. al. 2017).

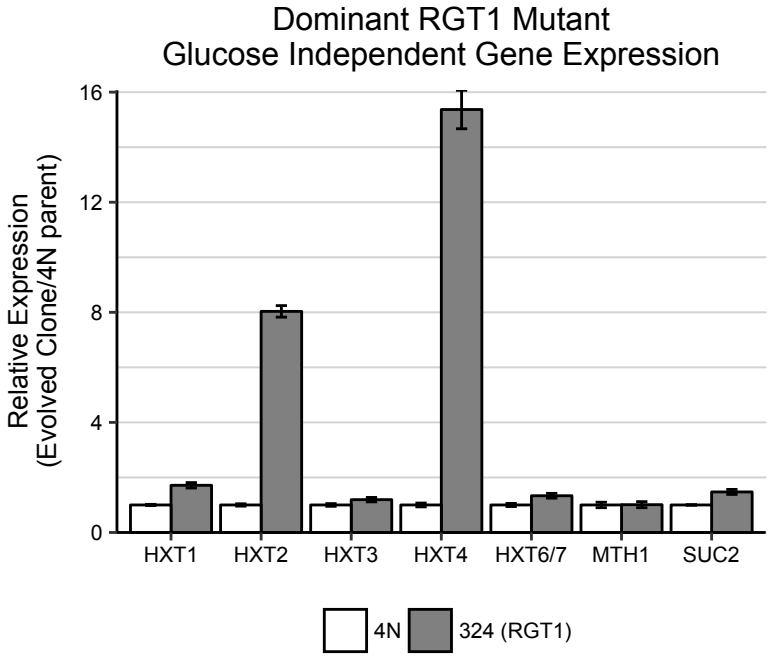


Figure 4-10 *RGT1-S509stop* induces glucose independent expression of *HXT2* and *HXT4*. qRT-PCR gene expression analysis of glucose responsive genes relative to *ACT1* expression and normalized to the 4N ancestral strain grown in 2% galactose. n=2, error bars represent SEM. Figure and legend adapted from (Scott et. al. 2017).

4.2.5 HXT6/7 amplification is more common at higher ploidy levels

Amplification of *HXT6/7* is a prominent pathway of adaptation to growth in glucose limiting conditions and we showed that *HXT6/7*^{amp} is more common with increasing ploidy (Figure 3-10) (Brown et al. 1998; Kao and Sherlock 2008; Kvitek and Sherlock 2011; Sellis et al. 2016). Only a single haploid in our study gained *HXT6/7*^{amp}, whereas other studies that adapted haploid yeast to growth in low-glucose conditions regularly recovered *HXT6/7*^{amp} mutants. This difference may be due to differences in culture conditions within our study. The majority of strains harboring *HXT6/7*^{amp} cluster together in clade *e* of the gene panel dendrogram. The two exceptions are 2Ne clone 207 and 1Ne clone 132, which also have mutations in *SNF3* and *IPT1,* respectively. Strains in *clade e* overexpress *HXT6/7* while down regulating the other hexose transporters (Figure 3-8). These strains also have reduced levels of the invertase *SUC2* compared to the diploid ancestor. This is surprising, given that Suc2p is required for the cells to metabolize raffinose.

Remarkably, we found that strains with *HXT6/7*^{amp} gained fewer mutations overall in the diploids ($p=1.92x10^{-3}$, paired student t-test) and tetraploids ($p=0.13$, not significant, paired student t-test) compared to strains with mutations in other genes, when adjusted for haploid genome content (Figure 4-11A). Despite the fact that tetraploid strains frequently underwent chromosome loss, and in some cases chromosomal gain, during adaptation to growth in raffinose (Chapter 2.2.1), strains that had *HXT6/7*^{amp} largely remained 4N. We compared propidium iodide staining between the strains with *HXT6/7*^{amp} to strains with other adaptive mutations and found that the *HXT6/7*^{amp} strains had significantly more DNA than the *SNF3* mutants ($p=2.96x10^{-3}$,

paired student t-test) and other mutants ($p=6.34x10^{-5}$, pair student t-test)(Figure 4-11B). These data suggest that amplification of the *HXT6/7* region has a genome stabilizing effect, potentially through reducing stress in the evolved clones (Chen et al. 2012; Oromendia et al. 2012; Shor et al. 2013). Further studies are required to understand the mechanisms of apparent genome stability in these strains.

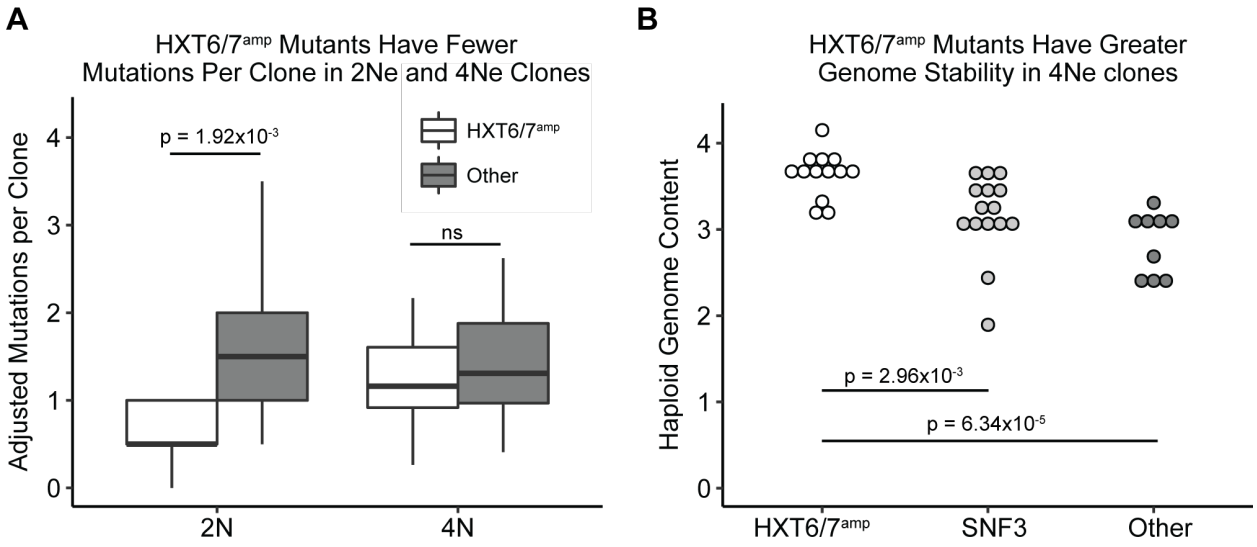


Figure 4-11 Evolved clones with amplification of the HXT6/7 region gain fewer mutations and remain primarily tetraploid. A) The average number of mutations per strain in the 2Ne clones and 4Ne clones with HXT6/7^{amp} (white) or without HXT6/7^{amp} (grey). The number of mutations per strain is adjusted for haploid genome content. The 2Ne clones with HXT6/7amp have significantly fewer mutations per strains than the clones with only a single copy of HXT6 and HXT7 ($p=1.92x10^{-3}$, paired student t-test). B) Haploid genome content (Y-axis), determined by mean G1 fluorescence, of tetraploid evolved clones with HXT6/7 amplifications, SNF3 mutations, or other mutations. Clones with HXT6/7^{amp} have a significantly larger haploid genome content than SNF3 mutants ($p=2.96x10^{-3}$, paired student t-test) and clones with other mutations ($p=6.34x10^{-5}$, paired student t-test). There is no significant difference between clones with mutations in SNF3 compared to clones with mutations in other genes.

4.2.6 Novel mutations in IPT1 and MOT3 improve fitness in raffinose media

The evolved haploid clones 103, 117, and 132 each gained unique non-synonymous or nonsense mutations in *IPT1*. Ipt1p, an inositolphosphotransferase, is the final step in the biosynthesis of complex sphingolipids, loss of which has been shown to improve antifungal drug resistance (Dickson et al. 1997; Hallstrom et al. 2001). It is improbable that the multiple mutations would occur in *IPT1* by chance ($p=6.2x10^{-6}$, exact binomial test), suggesting that mutations in *IPT1* are adaptive and under positive selection. The *IPT1* mutant strains slightly up-regulate *HXT3* and *HXT4* and cluster together in clade *d* (Figure 3-8). In a recent study that examined the fitness of gene deletions in several environments, loss of *IPT1* was predicted to have fitness benefits in low-glucose and low-sulfate conditions (Payen et al. 2016). However, ours is the first to report mutations in *IPT1* recovered from an experimental evolution.

Similarly, multiple clones (n=3) harbor mutations in *MOT3*, a transcription factor that has been shown to modulate the transcription of a wide range of genes, including ergosterol biosynthetic genes and glucose transporters (Grishin et al. 1998; Hongay et al. 2002). Overexpression of *MOT3* was shown to increase the transcription of *HXT2, HXT3, HXT4,* and *SUC2* (Grishin et al. 1998). As with *IPT1*, multiple hits in *MOT3* are unlikely to occur by chance ($p=5.0x10^{-6}$, exact bionomial test). The 1Ne clones 113, 114, and 120 have distinct frameshift (P268fs) and nonsense (R357Stop, K394Stop) mutations in *MOT3* (Table 9-1). These strains cluster together in clade *e,* adjacent to strains with $HXT6/7^{amp}$ and are characterized by up-regulation of *HXT3* and *HXT6/7* and down regulation of *HXT2, HXT4,* and *SUC2* (Figure 3-8).

In carbon limited environments, Mot3p forms a prion protein that represses Mot3p activity and the prion [*MOT3*$^{+}$] functions to prime yeast cells for growth in low-carbon environments by switching from fermentative to oxidative metabolism and producing a *FLO11*-induced multicellular phenotype in some lineages (Holmes et al. 2013). However, our background strain (S288c) does not express *FLO11*, thus it is unclear what role [*MOT3*$^{+}$] has in adaptation to growth in raffinose (Liu et al. 1996). The evolved clones with mutations in *MOT3* do not show evidence of prion formation by growth on glycerol glucosamine medium or multicellularity phenotypes (data not shown) (Halfmann et al. 2012; Holmes et al. 2013). Thus, it is unlikely that the *MOT3* mutants are adapting through prion formation.

Despite the unknown role of Ipt1p and Mot3p in adaptation to growth in raffinose, these strains exhibit high levels of competitive fitness compared to the diploid ancestor (Figure 4-1). One commonality between Ipt1p and Mot3p is they both function in the biosynthesis of plasma membrane constituents, specifically sphingolipids and ergosterol, respectively. This suggests that one mode of adaptation to nutrient limitation involves altering the plasma membrane or the endocytic pathway, including improved membrane organization, trafficking, recycling, and/or retention of the hexose transporters at the cell surface. In support of this, adaptive mutations were recently identified in genes encoding components of the endocytic pathway (including *FAB1* and *VAC14*) in yeast that had adapted to nitrogen limitation (Hong and Gresham 2014). These data support that a limiting factor for growth under nutrient limitation is the recycling or physical packing of membrane transporters at the cell surface (Mable 2001; Krogerus et al. 2016).

4.2.7 Chromosome XIII aneuploidy

As discussed in Chapter 2.2.1, aneuploidy of chromosome XIII occurred significantly more often than any other chromosome (Figure 2-6). Specific aneuploidies have been shown to be beneficial in certain growth conditions (Selmecki et al. 2006; Rancati et al. 2008; Chen et al. 2012; Mulla et al. 2014). However, due to other background mutations in the evolved clones isolated in our study, we cannot determine the specific beneficial effects of chromosome XIII aneuploidy in raffinose media. Thus, to test whether chromosome XIII gain contributed directly to the rapid adaptation of 4N cells, we generated isogenic 2N and 4N strains, with and without an extra copy of chromosome XIII.

The increased copy number of chromosome XIII provided a significant fitness increase to 4N strains in raffinose medium relative to the 2N ancestor (Figure 4-12A; t-test, $p<1x10^{-4}$) and there was no fitness benefit to chromosome XIII aneuploidy when grown in glucose (Figure 4-12B). This was not a general effect of aneuploidy because the gain of a different chromosome, chromosome XII, had the opposite effect on fitness (Figure 4-12A). In striking contrast to 4N cells, chromosome XIII trisomy was not beneficial to 2N strains in raffinose medium and decreased fitness of 2N cells in glucose. To our knowledge, this is the first observation of a ploidy-specific fitness advantage for an aneuploid chromosome. Thus, aneuploidy, acquired through high rates of mitotic errors, is one way that 4N cells can acquire more beneficial mutations with higher fitness effects.

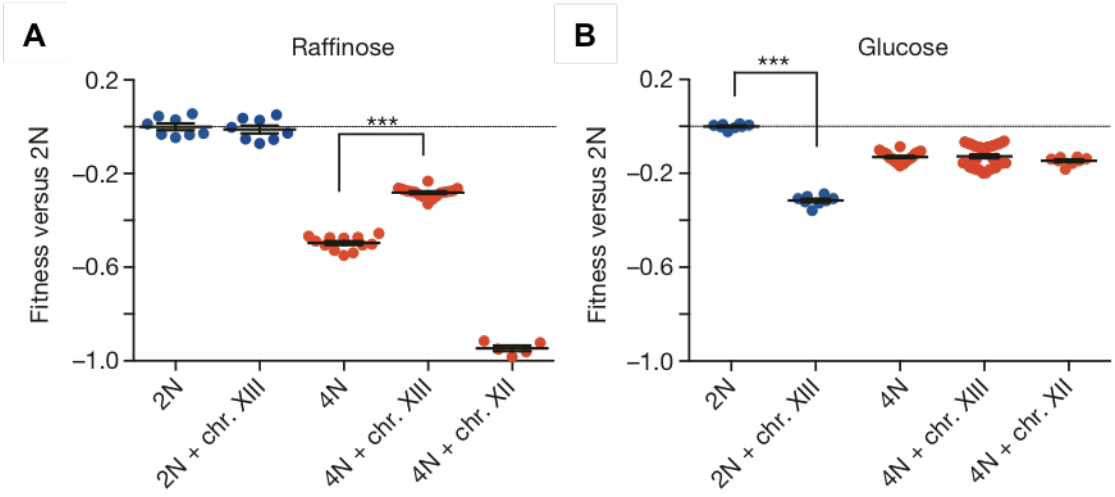


Figure 4-12 Chromosome XIII aneuploidy confers a raffinose specific benefit in tetraploid yeast. Gain of chromosome (chr.) XIII is beneficial to tetraploid cells grown in raffinose medium but not for diploids. Shown is the fitness of isogenic wild-type 2N and 4N strains, with or without chromosome XIII gain, relative to the 2N ancestor in A) raffinose or B) glucose medium. Error bars, mean +/- SEM of four individual clones and two or three technical replicates. Figure and legend adapted from (Selmecki et al. 2015).

4.3 Materials and Methods

4.3.1 Relative Fitness Assay

Anna Selmecki conducted the competitive fitness assays. Competitive fitness assays were performed using single colony isolates from the evolved populations and the diploid ancestral strain. One single colony was isolated from frozen stocks of each well of the evolution experiments (1N(A), 1N(B), 2N(A), 2N(B), 4N(A), 4N(B), 4N(C)) at generation 250. The evolved clones were cultured for 24h in 500 ml of Synthetic Complete + 2% raffinose, diluted into fresh medium, and competed with the ancestor expressing the complementary fluorescent protein. Competitions were initially performed using approximately the same number of cells from the ancestor and the evolved clone, but because the evolved clones grew significantly faster than the ancestor strains, the competitions were repeated using approximately five times more ancestor cells than evolved clone cells, with an initial population size of $1x10^5$. Serial dilutions were performed each day and the YFP:CFP ratio was determined by flow cytometry, yielding an estimate of the number of evolved (N_{t1}) cells relative to the ancestors (N_{t0}) as a function of time. The data were analyzed in Matlab using a custom script that performed a linear least-squares fit of $\log(N_{t1}/N_{t0})$ over multiple dilution cycles. The fitness relative to the ancestor is defined as s 5 d/dt $[\log_2(N_{t1}/N_{t0})]$, where t is measured in days (Hegreness 2006).

4.3.2 Strain construction

SNF3 mutant ploidy series. The *SNF3–G439E* mutation was constructed by Anna Selmecki in the haploid YFP strain back- ground (PY5999) using the pCORE counter-selectable reporter system (Storici et al. 2001), a gift from M. Resnick. Primers

SNF3_pCORE_KAN (5'-TGTTGGGGGTGTTATCATGACTATAGCCAACTTTATTGTGGCCATTGTTGGGAGCTCGTTTTCGACACTGG-3') and SNF3_pCORE_URA (5'-TATAAATGCTATCATAACTTTTGCGGCCGCTACAGTCTTTAAGGAACACTCCTTACCATTAAGTTGATC-3') were designed to integrate the CORE sequence at the SNF3 locus; PCR amplification and transformation procedures were followed as detailed previously (Storici and Resnick 2006). Sanger sequencing was used to identify clones with the desired mutation (chromosome IV: 112,896 G->A). Diploid *SNF3–G439E* mutants (heterozygous *SNF3–G439E/SNF3* and homozygous *SNF3–G439E/SNF3–G439E* clones) were constructed by mating after introduction of plasmids to confer mating competence (PB2649 or PB2647), as described in Chapter 8.1 for the construction of the CFP- and YFP-marked strains. An analogous strategy was used to generate tetraploid *SNF3–G439E* strains (heterozygous *SNF3–G439E/SNF3/SNF3/SNF3*).

Chromosome XIII ploidy series. The chromosome XIII aneuploid strain series was constructed by Anna Selmecki in the S288c background from the diploid strain PY7295 (RL4737) and the diploid PY7296 (RL4888), which is trisomic for chromosome XIII (Pavelka, Rancati, Zhu, et al. 2010). PY7296 was isolated from a triploid meiosis and a minimal number of cell divisions (Pavelka, Rancati, Zhu, et al. 2010). We confirmed the chromosome XIII trisomy by aCGH. We generated tetraploid clones by mating PY7295 to PY7296, with changes in mating-type accomplished as described previously (Storchová et al. 2006). Tetraploid clones were isolated on selective media and analyzed by flow cytometry and aCGH.

Rgt1 deletion strains. The *rgt1Δ* strain was produced by Amber Scott by replacing the entire *RGT1* ORF in the 5999 (1N) ancestral strain with the pCORE cassette (Storici et al. 2001). Transformants carrying the pCORE cassette were selected on YPD+200 μ g/mL G418 (GoldBio), and 4 clones were selected for confirmation. The *rgt1Δ* clones were grown to saturation in YPD+G418 and confirmed with PCR of *RGT1* region. A complete list of primers can be found in the Chapter 14.

4.3.3 Growth Conditions and Spot Assay

For glucose independent gene expression, cultures were grown to saturation in SC+2% raffinose media, cells were then diluted to 0.2 OD in SC+2% galactose and harvested after growth for 6 hours at 30°C. Spot assays were performed by growing strains to saturation in YPD or the corresponding selective media for *rgt1Δ* (YPD + 200μg/mL G418) or 4Ne-324 (SC+2% raffinose). Cells were washed and resuspended in water to a density of 1 ODU/ml. Cells were 10-fold serially diluted in water and 10ul was spotted on SC plates containing either 2% glucose or 2% raffinose. Plates were incubated at 30°C for 72 hours before photographing.

4.3.4 Expression analysis

RNA isolation, cDNA synthesis, and qRT-PCR was performed as described in Chapter 3.3. A list of qPCR primers can be found in Table 14-1.

4.4 Conclusions

To fully understand how organisms adapt to new or stressful growth environments, a detailed understanding of the ways in which evolution can act on the molecular level is critical. While previous studies have recovered adaptive mutations in some of the genes I characterize here, those studies largely ignored the effect of the individual mutations on gene function, organismal fitness and gene expression (Gresham et al. 2008; Kao and Sherlock 2008; Koschwanez et al. 2013; Kvitek and Sherlock 2013). In one study, mutations in *SNF3* were described as "non-tolerated missense mutations" (Kvitek and Sherlock 2013). However, as I demonstrated, mutations in *SNF3* result in constitutive de-repression of the hexose transporters. Similarly, multiple studies have reported loss-of-function mutations in *RGT1* (Koschwanez et al. 2013; Kvitek and Sherlock 2013), yet I show that there is no benefit to loss of *RGT1* in glucose-limited growth and the mutations previously reported in *RGT1* are all located in the same "inhibitor of repression" domain, strongly suggesting that these mutations may, in fact, be dominant.

In this chapter, I characterized mutations recovered at all ploidy levels. We found novel mutations in haploid populations in *IPT1* and *MOT3* that suggest these genes may have a previously unknown role in the glucose pathway. We hypothesize that adaptation through mutations in *IPT1* and *MOT3* maybe alter the composition of cell wall constituents, such as glucose transporters, though the endocytic pathway. We also discover novel dominant mutations in *RGT1* and *MTH1* in tetraploid evolved populations. Additionally, chromosome XIII aneuploidy is the first example of a polyploid specific beneficial mutation, as it was only beneficial in the tetraploid

background in raffinose. On the other hand, constitutive activating mutations in the glucose sensors *SNF3* and *RGT1* were found all ploidy levels. *HXT6/7*^{amp} was also discovered at all ploidy levels, but increased in frequency with increasing ploidy level. Surprisingly, *HXT6/7*^{amp} strains exhibit fewer mutations and larger genome size, consistent with increased genome stability. Thus, using organisms of multiple ploidy levels in our studies allowed us to discover novel dominant mutations and fully characterize the effects of individual mutations at ploidy levels relevant to nature (Otto 2007).

5 CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Summary

This study is the first of its kind to integrate both gene expression and whole genome sequencing on a large scale, particularly with different levels of cellular ploidy. Whole genome sequencing with variant calling was performed on over 100 evolved clones isolated from individual evolving populations. We measured the expression profiles of genes shown to be indicative of adaptation to growth in raffinose media in each evolved clone. For the majority of the evolved clones, we further determined the specific adaptive mutation responsible for the expression profile. Finally, we measured the relative fitness gains for each evolved clone in raffinose media. This is the first study to quantify and describe genome variants, measure gene expression and estimate fitness in haploid, diploid and, in particular, tetraploid evolved clones.

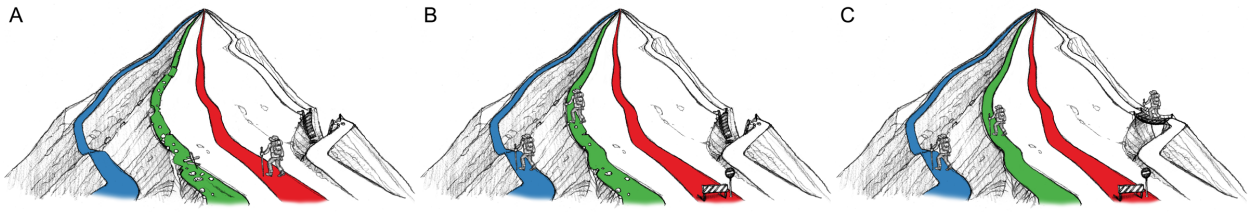


Figure 5-1 Graphical summary of pathways of adaptation for haploid, diploid, and tetraploid yeast to growth in raffinose media. Cartoon depicting the differing pathways to adaptation to long-term growth in raffinose media by the A) haploid, B) diploid, and C) tetraploid evolved clones. Non-synonymous mutations, such as *SNF3*, are represented by the blue path. Copy number variations, such as *HXT6/7*^{amp}, are represented by the green path. Recessive mutations, such as *MTH1*, are represented by the red path. Aneuploidy, only observed in the 4Ne populations, is represented by the white pathway. The hiker signifies the commonly utilized pathways in the clones at each ploidy level. Artwork courtesy of David Deen. Figure and legend from (Scott et. al. 2017)

We find that underlying ploidy level has a major effect on the spectrum of mutations gained during adaptation. The observed difference in types of mutations acquired by the evolved clones is reflected in the different pathways of adaptation to growth in raffinose (Figure 5-1). Not all pathways of adaptation are available to strains from different ploidy level backgrounds. For example, the most common pathway of adaptation in the evolved haploid clones is through recessive mutation, which is inaccessible in the diploids and tetraploids. Similarly, adaptation through aneuploidy occurred exclusively in the evolved tetraploids, as previously described (Selmecki et al. 2015). While copy number variation is rare in the haploid evolved clones, it is more accessible with increasing ploidy level (Gresham et al. 2008; Sellis et al. 2016). Some adaptive routes are open to all ploidy levels: we see non-synonymous mutations in the glucose sensors in haploid, diploid, and tetraploid evolved clones. This study demonstrates the adaptive potential of higher ploidy organisms during evolution and shows that different evolutionary outcomes arise as a result of ploidy level (Schoustra et al. 2007).

Historically, experimental microbial evolution studies were performed exclusively in haploid organisms, and in more recent years diploid organisms. This study establishes that using a range of ploidy levels in evolution experiments is necessary to fully understand the mechanisms of evolution, given the prevalence of whole genome duplication events throughout evolutionary history. Utilizing a range of ploidy levels in our study allowed us to identify novel mutations, especially mutations that are physiologically relevant to diploids and polyploid organisms.

5.2 Major Scientific Contributions

In this study we demonstrate that polyploidy may drive rapid adaptation in yeast by various mechanisms. We show that polyploidy increases the genetic diversity of the evolving populations. Tetraploid yeast acquire a greater number of mutations and gain more diverse adaptive mutations than haploids and diploids, indicative of greater flexibility during adaptation. If these mutations are beneficial at lower ploidy states, then the long-term benefit of polyploidy will be preserved, even if polyploidy is transient during adaptation. Our study provides strong evidence for a beneficial role of ancient polyploidization events that occurred in the evolution of most species.

To our knowledge, this is the first study to identify examples of mutations that are selectively beneficial in polyploid strains. Whole chromosome aneuploidy of chromosome XIII provided a tetraploid specific benefit in raffinose media. Additionally, an SNF3 mutation was more beneficial in a tetraploid background than haploid or diploid backgrounds. We are also the first study to suggest that a particular adaptive mutation, *HXT6/7*^{amp}, may have a genome stabilizing effect in polyploid yeast.

Despite it being a well-studied and characterized pathway, we also make significant discoveries relevant to glucose pathways in yeast. We found adaptive mutations in two genes, *IPT1* and *MOT3,* with no known role in the canonical glucose sensing and signaling pathways. Strains harboring mutations in these genes exhibit impressive fitness gains in raffinose media (Figure 4-1). By studying these mutations further, we may discover more about the complex regulation of nutrient metabolism in yeast. We also discover novel dominant mutations in *MTH1* and *RGT1.* These mutations can be used to better understand the precise function of these genes in

regulating glucose responsive gene expression. The glucose pathway in yeast shares many of the same factors as the glucose pathway in humans (Ozcan and Johnston 1995), thus contributions to the field of regulating glucose uptake in yeast may lead to a better understanding of glucose sensing and transport in mammals.

5.3 Future Directions

5.3.1 Evolutionary studies

There remains a wealth of information to be gleaned from our experimental evolution study. In my thesis I focus primarily on single clones isolated from each population at a single time-point, generation 250. However, frozen stocks of the evolution experiments were made every 3-4 days throughout the experiment. These stocks allow us to ask several questions pertaining to the process of adaptive evolution. In particular, I would like to sequence single evolving populations at many time points to look at the order in which adaptive mutations are gained, how quickly those mutations spread, and the identity of other mutations that exist transiently within the population. How does the starting ploidy influence these characteristics? We could ask if there were certain types of mutations (like copy number variation) occurring relatively earlier in evolution (Gresham et al. 2008). Are chromosomes gained or lost in pairs, as is suggested by our pairwise analysis (Figure 2-5)? We find a greater sampling of adaptive mutations in the tetraploid evolved clones, is this reflective of greater population diversity? We can address this question by examining he diversity of whole populations (rather than the single isolated clone utilized here), especially the tetraploid populations.

Additionally, we evolved a subset of tetraploid clones an additional 250 generations (denoted as 4Ne500). We have sequenced a subset of these strains, but

otherwise are only beginning to examine this population. What we know so far is that nearly all the 4Ne500 strains have diploidized, although there are a few exceptions, in which the 4Ne500 clones have remained 4N. This result is consistent with other studies that indicate that polyploidy is a transient state (Gerstein et al. 2006; Voordeckers et al. 2015). Studies have shown that early beneficial mutations have the greatest fitness effects and subsequent mutations continue to improve fitness, though to a lesser degree (Orr 2005). In the 4Ne500 clones we could ask what types of mutations are subsequently gained and what effect do these mutations have on overall fitness. We could also compare 4Ne500 clones to their generation 250 ancestors. For example, do the 4Ne500 clones exhibit greater trade-offs in other growth conditions than 4Ne250 clones? Have they become more specialized to growth in raffinose or, as has been previously reported, do subsequent adaptive mutations improve growth independent of carbon source (Wenger et al. 2011)? Overall, there is much more that can be learned about the pathways of adaptation, their relative fitness, and their frequency by studying the strains we already have in the freezer.

5.3.2 Characterization of novel mutations

IPT1* and *MOT3 The first step in further understanding the IPT1 and MOT3 mutations is to confirm conclusively that these mutations are indeed causally responsible for the adaptive phenotype. Hence the first step would be to create strains containing only these mutations and confirm that they indeed confer improved growth in raffinose relative to the ancestor. Given that we observe multiple clones with these mutations, this is expected to be purely a confirmatory experiment.

Once confirmed, the next step to understanding the adaptive mechanisms of the *IPT1* and *MOT3* mutations is to test whether or not these strains adapted to growth in raffinose by importing more glucose into the cell (as seen in the other adaptive mutants). Thus, I would begin by measuring the concentration of glucose in the media over time (Roop et al. 2016). I would also measure the population density over time, as some mutations are thought to impact growth in lag-phase and stationary-phase, rather than improve the growth rate (Toussaint and Conconi 2006; New et al. 2014). A shortened lag-phase or a higher density stationary-phase may explain why studies of chemostatic growth have not recovered mutations in these genes, as growth in a chemostat maintains constant population size whereas in serial transfer experiments, cells undergo daily lag- and stationary-phase growth.

We observe only modest expression changes in the glucose responsive gene panel strains with mutations in *IPT1* and *MOT3*. Thus, these mutations may be working post translationally to alter protein concentrations or localization. To test this, I would examine the localization of transporters by fluorescent microscopy (since there is no evidence of increased signaling from the sensors). With this strategy, I believe we would begin to understand the role of *IPT1* and *MOT3* in adaptation to growth in raffinose.

Dominant *RGT1* and *MTH1* mutations As with other mutations, the first step is to confirm that these dominant mutations are indeed causal for the adaptive mutation. To this end, I have made efforts to engineer strains that harbor only the MTH1-C321F or RGT1-S509stop mutation, free from other, presumably hitchhiker mutations observed in the generation 250 adaptive clones. These strains can be used to determine the fitness

effect of each dominant mutation in differing ploidy backgrounds, similar to the SNF3-G439E mutation (Figure 4-4).

Surprisingly, the genome wide binding profile of Rgt1p has never been measured. Experiments that look for changes in the Rgt1p binding profile under specific conditions, such as low- and high- glucose media and galactose media, would go a long way in elucidating the different functional roles of Rgt1p in different environments. My work indicates that the *RGT1-S509Stop* mutation is not a complete loss-of-function (Chapter 4.2.4); therefore, performing ChIP-seq or targeted ChIP-qPCR on this truncated form on Rgt1p may be one method to understand the molecular effects of the Rgt1p "inhibitor of repression" domain (Figure 4-8) (Polish et al. 2005). To this end, I have ongoing efforts to see if the truncated form of Rgt1p is translated, thanks to antibodies to both the C-terminus and N-terminus of Rgt1p, kindly provided by Mark Johnston. I am further working to optimize these antibodies for use in ChIP-qPCR and ChIP-seq.

Glucose independent expression in the *RGT1-S509Stop* mutant strain suggests that the mutation may only impact Rgt1p function at specific promoters, specifically *HXT2* and *HXT4* (Figure 4-10). I hypothesize that this may be the result of dissociation with a specific binding partner at those promoters. Therefore, another method to dissect the role of dominant mutations in *MTH1* and *RGT1* would be to look at their association with known regulators and binding partners such as Ssn6p/Tup1p and Std1p (Gancedo 2008) or measure protein interactions en masse with mass spectrometry.

5.3.3 Tetraploidy induced environmental stress response

The environmental stress response in yeast is characterized by rapid, but transient, transcriptional changes in a stereotypical set of genes after introduction of the stressor. The expression changes typically spike at 15 minutes after introduction of the stress and return to basal levels within an hour (Gasch and Werner-Washburne 2002). Yet, the tetraploid ancestor is exhibiting signs of ESR after 24 hours of growth in raffinose (Figure 3-2), albeit with considerable variability (Figure 3-3). To test the impact of ploidy on the ESR immediately upon transfer to raffinose media, an undergraduate in the lab, Hannah Chatwin, collected RNA at 15min, 30min, 1hr, 1.5hr, and 2hr after switching log-phase cells from rich media (YPD) to either rich media or raffinose media. She tested four known ESR genes and their non-ESR iso-enzymes, a representative plot is shown in Figure 5-2. Her results indicate that there is no noticeable difference in the initial environmental stress response in tetraploid yeast when grown in raffinose. If this early result holds, it would suggest that cells of all ploidy levels respond similarly at the onset of stress. However, the effect of long-term stress (as indicated by the RNA-seq and qRT-PCR) on mutability is relatively unknown. A great deal of additional work is necessary to fully understand the temporal dynamics of activation of ESR in the tetraploid ancestor in raffinose media.

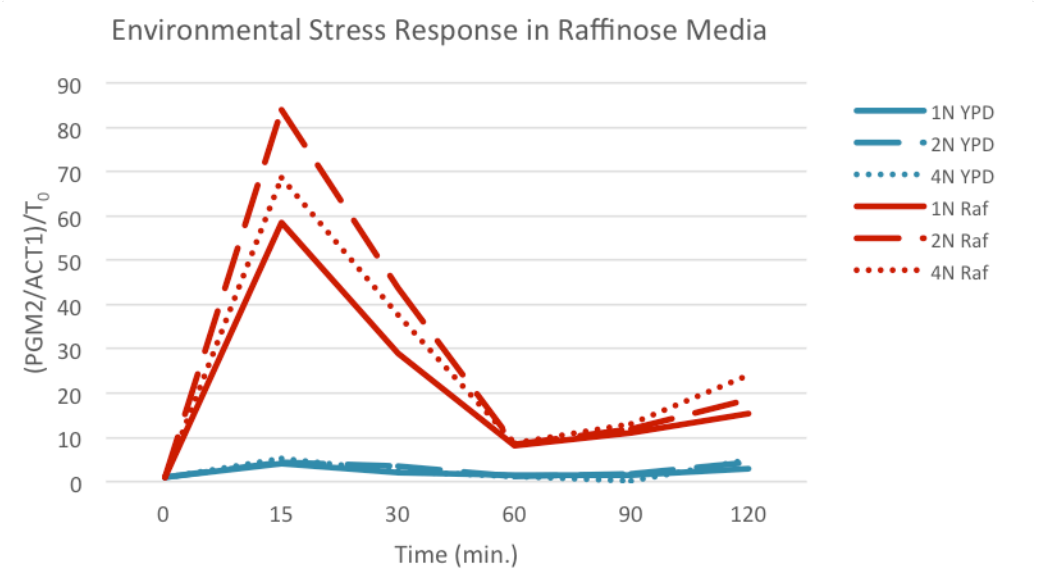


Figure 5-2 Ploidy level does not impact the environmental stress response to growth in raffinose media The 1N (solid), 2N (dashed), and 4N (dotted) ancestral strains were grown to log-phase in YPD. At time zero, strains were switched to either YPD (red) or raffinose media (blue). Expression of the ESR gene PGM2 peaks at 15minutes and returns to near baseline after 1 hour. Ploidy does not impact the initial ESR when strains are introduced to raffinose media. Expression of *PGM2* was measured by qRT-PCR and normalized to *ACT1* expression. All values are relative to the diploid ancestral strain at T_0. n=1.

5.3.4 Genome stabilization by HXT6/7^{amp}

I think one of the most interesting, and potentially most controversial discoveries that I made was the apparent genome stabilizing effect of *HXT6/7*^{amp}. Strains that harbor *HXT6/7*^{amp} gained significantly fewer mutations and remained largely tetraploid (Figure 4-11). I hypothesize that the stabilizing effect is related to a reduction in stress in the tetraploid lineages that gained amplifications of *HXT6/7*. It has been shown that polyploidy and aneuploidy can occur in response to stress (Storchova and Pellman 2004). In fact, aneuploidy itself causes a general proteotoxic stress response (Oromendia et al. 2012). Our whole genome expression data suggest that the tetraploid

ancestor is more stressed in raffinose than the diploid ancestor (Figure 3-2), thus, it is not unreasonable that reducing stress could result in genome stabilization.

To further understand the impact of *HXT6/7*^{amp} on the environmental stress response, we compared the diploid evolved clones with mutations in *SNF3* and *HXT6/7*^{amp} (2Ne clones 232 and 233, respectively) to the diploid ancestor. Whole genome expression in the strain harboring *HXT6/7*^{amp} was negatively enriched for ESR genes, suggesting it was indeed less stressed than the diploid ancestor (data not shown). In contrast, the strain with an *SNF3* mutation showed no significant alteration in the stress response, relative to the diploid ancestor. While these data suggest that *HXT6/7*^{amp} may reduce stress relative to *SNF3*, further studies are required to confirm these preliminary results. If *HXT6/7*^{amp} does reduce stress, further studies are needed to understand how its amplification reduces cellular stress.

While the evidence I present in chapter 4.2.5 is suggestive of a genome stabilizing effect of *HXT6/7*^{amp}, direct measurements of *HXT6/7*^{amp} on genome stability are required. One method is to repeat the evolution experiment with strains engineered to have either *HXT6/7*^{amp} or another adaptive mutant, such as *SNF3*. Genome stability could be measured quickly by quantifying total DNA content over time, as previously described (Gerstein et al. 2006). If *HXT6/7*^{amp} is genome stabilizing, I hypothesize that strains with *HXT6/7*^{amp} will remain largely tetraploid while strains with other adaptive mutations would be expected to converge on diploidy.

6 THE IMPACT OF ANEUPLOIDY ON TRANSCRIPTIONAL REGULATORS

6.1 Introduction

In addition to my primary research project, I also sought to understand the impact of aneuploid on transcriptional regulators. Aneuploidy, when cells posses a karyotype that is not a multiple of the haploid complement, has a profound impact on cellular functions (Torres et al. 2008). In our experimental evolutions study of yeast strains, we observe that most tetraploid strains become aneuploid over the course of adaptation. In terms of human health, 10-50% of human embryos are aneuploidy (Nagaoka et al. 2012), it is a common cause of human miscarriage and malformations (Fitzpatrick 2005), the leading cause of intellectual disability (Oromendia and Amon 2014), is linked to premature aging (Wijshake et al. 2012), and is associated with many cancers (Weaver and Cleveland 2006; Nagaoka et al. 2012; Chen et al. 2015). Transcriptional regulation is governed by the interaction of a large set of DNA-binding proteins including transcription factors, nucleosomes, histone modifying enzymes, and basal transcriptional machinery. The binding specificity of many of these proteins comes from both their specific sequence affinity and cellular abundance. Furthermore, their effect on target gene expression is influenced by interactions with other DNA-binding proteins, an effect often referred to as combinatorial control. An interesting, non-obvious, and yet critically important characteristic of the regulatory network is that small changes at key points within the network can result in dramatic changes in the resultant expression profile. It stands to reason that changes in DNA copy number due to aneuploidy has major effects on the gene regulatory network through DNA binding proteins.

Down syndrome (T21) is a common human autosomal aneuploidy caused by an

extra copy of chromosome 21 (HSA21). Individuals with T21 exhibit heart defects (Li et al. 2012), cognitive and developmental deficits (Lott and Dierssen 2010; Dierssen 2012), and an increased risk of Alzheimer's disease, childhood leukemia, and other autoimmune diseases (Yoshida et al. 2013; Hartley et al. 2015; Colvin and Yeager 2017). However, the spectrum and severity of phenotypes varies greatly between individuals (Potter 2016). Ultimately, by understanding the molecular basis of Down syndrome, particularly on transcription, we gain key insights into all aneuploidy-associated aspects of biology.

Most studies suggest that the Down syndrome phenotypes are caused by a gene dosage imbalance for genes encoded on HSA21, including transcription factors (TFs) and constituents of protein complexes (Gardiner 2006; Megarbane et al. 2009). It has been shown that an increase in chromosome copy number results in a concomitant increase in RNA and protein levels, known as the primary dosage response (Figure 6-1)(Torres et al. 2010; Tang and Amon 2013). However, there are several genes located off HSA21 that also exhibit alterations in expression (Fitzpatrick 2005; Costa et al. 2011; Vilardell et al. 2011; Sheltzer et al. 2012). These alterations off of HSA21 may be the results of trans-acting genes dosages effects (Fitzpatrick 2005; Antonarakis 2016). For example, transcription factors encoded on HSA21 may alter the expression of genes on other chromosomes.

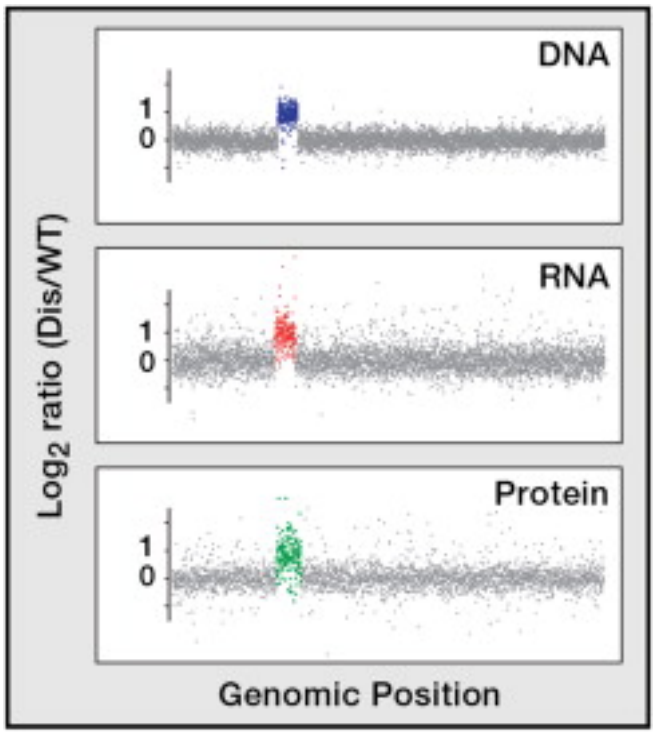


Figure 6-1 Aneuploidy increases DNA, RNA, and protein levels on altered chromosome. In yeast disomes, RNA and protein levels increase with the increased DNA copy number of aneuploid chromosomes. Figure is used with permission from (Tang and Amon 2013).

Indeed, a meta-analysis of multiple expression studies indicate genes involved in transcriptional regulation, such as transcription factors, are enriched in the dysregulated genes in Down syndrome (Vilardell et al. 2011). Additionally, HSA21 itself contains several TFs that are altered in expression in individuals with Down syndrome (Prandini et al. 2007). Studies have shown that overexpression of the HSA21 TFs Oligo2 and Ets2 in mouse cells disrupt neurogenesis and cause neuronal apoptosis similar to that in Down syndrome brains (Wolvetang et al. 2003; Liu et al. 2015). While these studies suggest these phenotypes are a result of downstream gene misregulation, they do not directly assess the mechanism of the dysregulation. To understand the molecular

mechanism underlying in impact of trisomy 21 on the transcriptional regulatory network it is necessary to directly profile DNA-binding proteins genome wide in both Down syndrome and euploid cells.

My long-term goal of this project is to understanding the impact of aneuploidy, specifically trisomy 21, on DNA-binding proteins. I hypothesize that changes in the DNA/transcription factor stoichiometry due to trisomy 21 results in aberrant transcription factor binding (Figure 6-2). In particular, I hypothesize that transcription factors encoded on HSA21 with increased expression will have an altered genome-wide binding profile: they may exhibit an increased binding frequency at known binding locations and/or bind lesser affinity binding sites. For these studies I am utilizing lymphoblastoid cell lines (LCLs) derived from a family of individuals: a mother, father, child with Down syndrome, and an unaffected brother (Figure 6-3). The brother is of a similar age as the child with Down syndrome and serves as an age and sex matched control. By performing these studies in the family, we can also see the impact of inheritance and individual variation on TF binding. In this chapter, I discuss preliminary experiments to determine the impact of T21 on the genome-wide binding profiles of TFs encoded on HSA21.

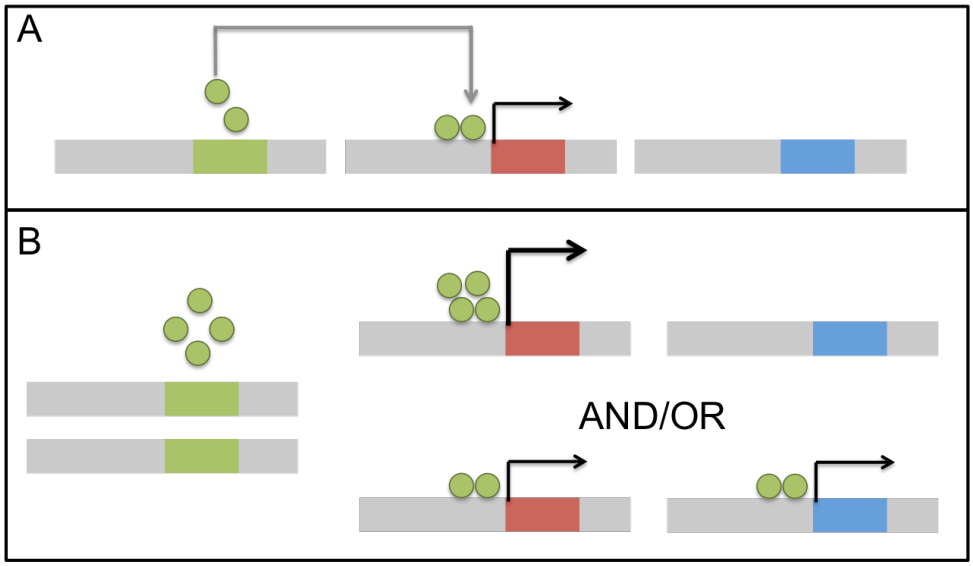


Figure 6-2 Graphical hypothesis of the impact of transcription factor copy number on genome-wide binding profile. A) Example of transcription factor binding in euploid cells. The TF (green) typically binds to the promoter of the red gene and modulates transcription. B) Example of altered transcription factor binding in cells with increased TF copy number. Increased copy number of the TF (green) results in increased TF protein production. This may lead to increased frequency of binding at the red gene, or off target binding at other sites, such as the blue gene.

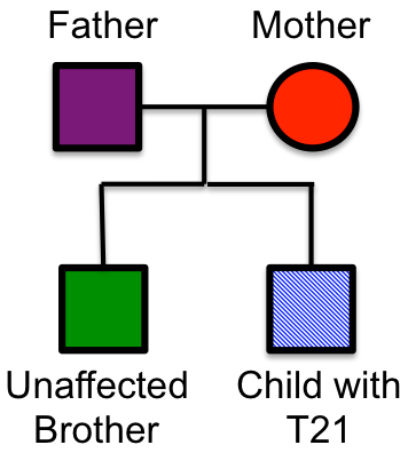


Figure 6-3 Pedigree for the lymphoblastoid cell lines used in this study. The cells used in this study were derived from a family with a mother (red), father (purple), a child with Down syndrome (blue), and unaffected brother (green). These colors are used in subsequent figures to denote the different individuals in the family.

6.2 Results and Discussion

6.2.1 Expression of TFs encoded on HSA21 are increased in T21

There are several transcription factors encoded on HSA21, however not all of them are expressed in each cell type. To determine which HSA21 TFs are expressed in LCLs, I examined RNA sequencing performed by Dr. Mary Allen in the familial derived LCLs. I found that only a handful of HSA21 transcription factors were expressed in LCLs: GABPA, BACH1, RUNX1, and ETS2 (data not shown). I next asked if these TFs were upregulated, relative to euploid controls, in the cells derived from the individual with T21. I measured the expression of GABPA, BACH1, ETS2, and RUNX1 in each individual in the familial derived LCLs with RT-qPCR. Gene expression is internally normalized to RPL13 (Figure 6-4). The cells derived from the individual with T21 up-regulate the HSA21 encoded TFs ~1.5 fold compared to the cells derived from euploid individuals. These results are consistent with expression levels measured in LCLs and fibroblasts isolated from a larger cohort of T21 and euploid individuals (Prandini et al. 2007). Although, there is some variability in the expression levels of the TFs between the individuals measured, these results indicate that increased chromosome copy number due to T21 increases the expression of genes, particularly TFs, encoded on HSA21.

Importantly, each of these TFs have a known role in the T21 phenotype, suggesting that overexpression of these TFs due to an additionally copy of HSA21 may have functional consequences. BACH1 is involved in oxidative stress, cell cycle control, and neurodegeneration and is overexpressed in the brains of DS and Alzheimer's patients (Shim KS, Ferrando-Miguel R 2003; Warnatz et al. 2011). GABPA has been

shown to regulate mitochondrial function and play a role in the Down syndrome phenotype (Rosmarin 2004; Gardiner 2006). Lastly, RUNX1, which regulates hematopoiesis and ETS2 have been implicated in leukemia in T21 (Sementchenko and Watson 2000; Fonatsch 2010; Nižetić and Groet 2012).

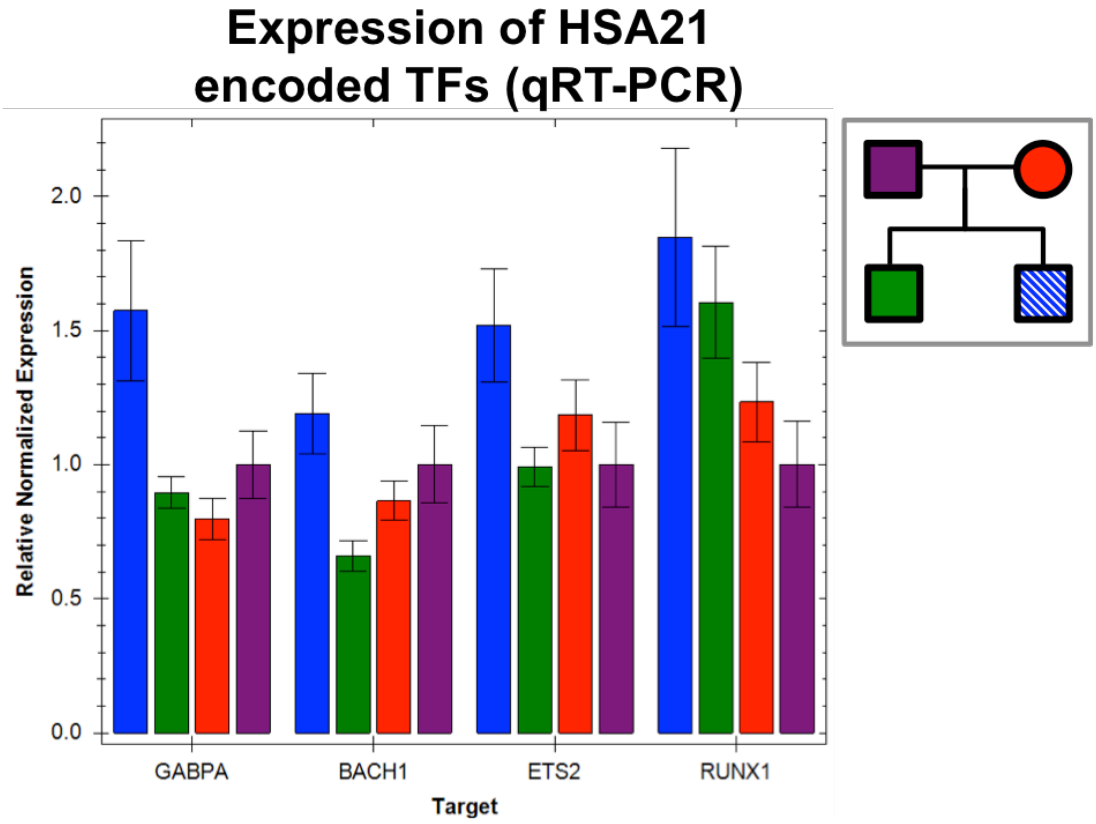


Figure 6-4 Expression of HSA21 encoded transcription factors in familial derived lymphoblastoid cell lines. Relative normalized expression for HSA21 encoded TFs in cells derived from an individual with T21 (blue), his mother (red), his father (purple) and an unaffected brother (green). Expression of the indicated TF (x-axis) was measured by RT-qPCR and internally normalized to RPL13 expression and relative to expression in the father (purple). Error bars represent the standard error of the mean for 2 biological replicates.

6.2.2 Lack of alterations to GABPA binding profile in T21

I hypothesized that increased TF dosage due to T21 would result in alterations to the genome-wide transcription factor binding profile of TFs encoded on HSA21. To test this hypothesis, I profiled GABPA binding in the familial derived LCLs. I selected GABPA for further analysis because the euploid cell lines expressed GABPA is similar levels and the T21 cell line overexpressed GABPA at 1.5 fold. Additionally, GABPA has been profiled in the ENCODE project in several cell types, included lymphoblastoid cells, thus there exists a well-characterized GABPA antibody available for chromatin immunoprecipitation (ChIP) (The ENCODE Project Consortium 2011). I performed ChIP-seq to determine the genome-wide binding profile for GABPA with mouse IgG as a control following ENCODE standards (Landt et al. 2012). Each library had between 16 to 24 million uniquely mapped reads (Table 6-3) and peak-calling analysis found between 1000 and 2100 peaks per cell line for a merged total of 2183 GABPA peaks (Table 6-1). I also confirmed that the cell line derived from the individual with T21 had an extra copy of HSA21 (Figure 6-5).

Cell line	Peaks	Unique Peaks
T21	1424	43
2N Brother	1041	7
2N Mom	1259	19
2N Dad	2107	609
Merged Total		**2183**

Table 6-1 Summary of GABPA Peak Calling

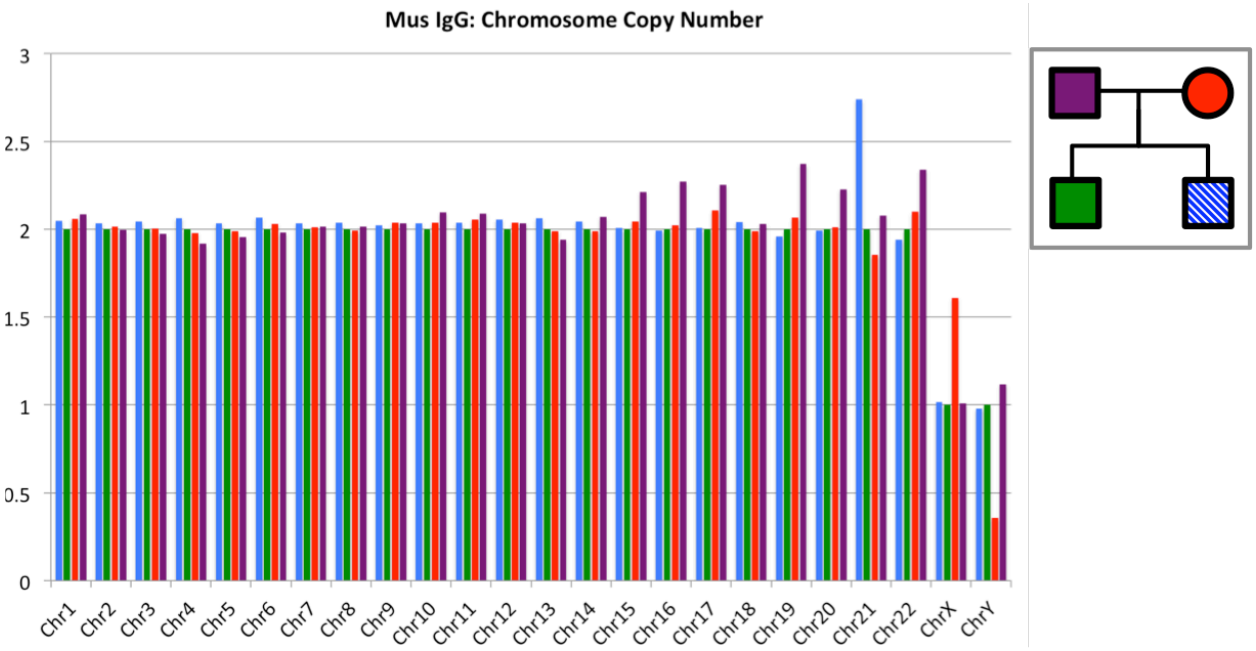


Figure 6-5 Chromosome copy number in ChIP-seq data Average number of sequencing reads per chromosome in the child with T21 (blue), 2N brother (green), 2N mother (red) and 2N father (purple). Chromosome copy number was determined from mouse IgG ChIP sequencing libraries normalized to coverage in the 2N brother (green).

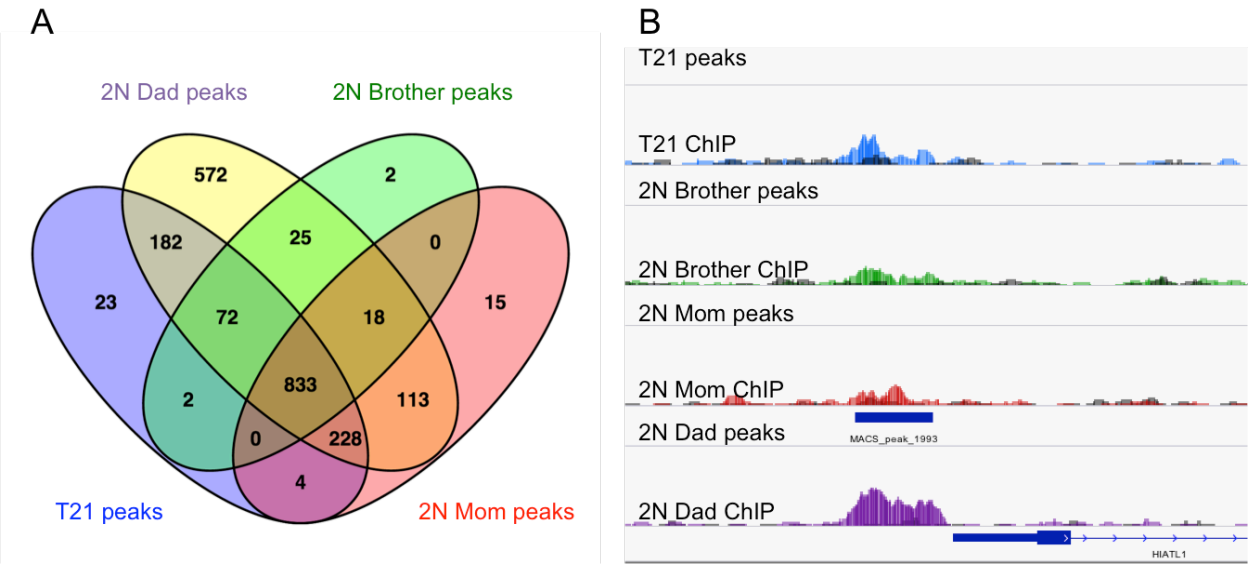


Figure 6-6 GABPA unique peaks in familial derived LCLs. A) Venn diagram comparing the shared and unique peaks in the familial derived LCLs. There are a greater number of shared and unique peaks called in the cells derived from the 2N Dad. Venn diagram created with Venny (Oliveros 2015). B) IGV snapshot of a peak region that was called unique in the 2N Dad cell line.

Using the ChIP-seq peak calls, I asked if GABPA was bound at different locations in each individual. In my initial analysis I asked which peaks overlapped between cell lines (e.g. “shared”) and which peaks were unique to each cell line (Figure 6-6A). Overall, the GABPA libraries prepared from the cells derived from the euploid father had a greater number of peaks. Unfortunately, in many cases when the peak-calling algorithm called peaks in only a single cell line, there appears to be coverage suggestive of lower level undetected peaks in the other lines as well (Figure 6-6B). To account for this issue and determine the unique peaks for each individual I instead merged all peaks identified in the ChIP-seq analysis (any individual) and then determined which of these peaks were present in each cell line above background (Figure 6-7). Background was determined to be one standard deviation above the median normalized coverage over the merged peak regions (FPKM) in the mouse IgG libraries (described in section 6.3.5). Overall, there were very few individual GABPA peaks identified between the familial derived LCLs, suggesting that increased gene dosage of GABPA does not increase GABPA binding sites genome-wide.

I next asked if there was increased occupancy of GABPA due to increased GABPA gene dosage in the individual with T21. To determine differential occupancy, I counted the number of reads over all GABPA binding sites identified in the ChIP-seq (any individual) and performed differential expression analysis using DESeq between pairs of individuals (Figure 6-8). There are at most 4 GABPA binding sites that exhibit differential occupancy between the individual with T21 and any other members of his family, indicating that ChIP occupancy is not dramatically altered for GABPA as a result of increased gene dosage due to T21.

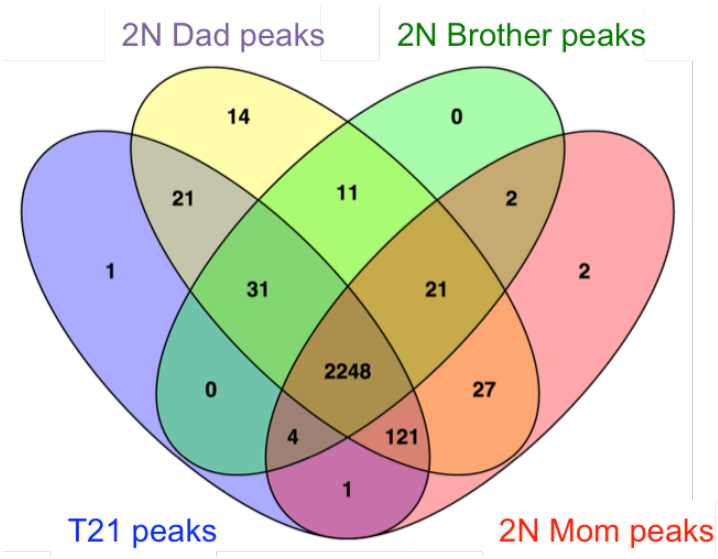


Figure 6-7 GABPA peaks in each individual with coverage above background. Venn diagram comparing the shared and unique peaks with coverage above background in the familial derived LCLs. The majority of peaks are present in each all cell lines. Venn diagram created with Venny (Oliveros 2015).

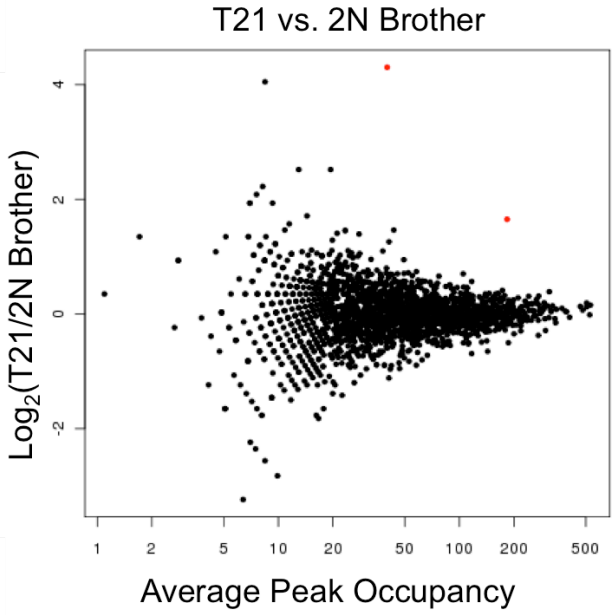


Figure 6-8 Differential GABPA peak occupancy. MA plot depicting the differential occupancy of GABPA peaks between cells derived from the individual with T21 and his euploid brother. Each dot is a single peak, the X-axis is average peak occupancy and the Y-axis is the $\log_2$ transformed differential occupancy between cells derived from the individual with T21 and his euploid brother. Red dots are significantly differential occupied sites.

Overall, there was little to no alterations to the genome-wide binding profile of GABPA in the individual with T21, despite the increased gene dosage identified by RT-qPCR (Figure 6-4). This may be due to several factors. First, these results are preliminary and only performed in a single biological replicate for each cell line, more replicates are required to draw significant conclusions. Additionally, while GABPA is over-expressed in the T21 cells on the RNA-level, protein levels of GABPA may be tightly regulated and thus dosage compensated. A preliminary analysis of GABPA protein levels in the LCLs derived from the individual with T21 and his brother do not show an increase in GABPA protein (Figure 6-9). Alternatively, ChIP-seq analysis may not be sensitive enough to detect small changes in GABPA binding in a population of cells. In particular, it is very difficult to normalized ChIP-seq data between samples. Even in well-controlled ChIP-seq experiments, variability in crosslinking efficiency, shearing, and antibody pull-down are expected. One method to improve normalization is to use spike-in controls, i.e. to add an equal number of cells from a different species into each sample before crosslinking and normalize the ChIP-seq libraries to the number of reads that map to the exogenous genome (Bonhoure et al. 2014; Orlando et al. 2014). However, this method requires an antibody that reacts to both human and exogenous cells. Ultimately, better normalization and an increased number of replicates are necessary to accurately measure small alterations to the genome-wide binding profiles of transcriptional regulators.

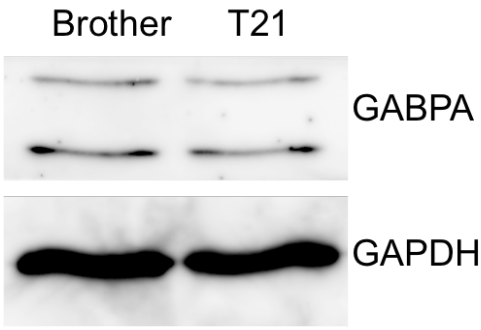


Figure 6-9 GABPA protein expression is not altered in T21 cells. Western blot for GABPA in T21 cells and euploid cells with GAPDH as a loading control.

6.2.3 Alterations in H3K4me3 binding are due to individual variation

We cannot know *a priori* which transcription factors would have a dosage sensitive response on genome-wide binding a gene regulation; therefore, I also examined general promoter usage between familial derived LCLs. To do this I performed ChIP-seq for H3K4me3, an epigenetic marker of promoters (Kouzarides 2007). There are an average of 25,000 peaks called in each individual (Table 6-2). I further compared the location of the H3K4me3 histone marks between the individual LCLs (Figure 6-10). The majority of H3K4me3 sites are shared between cells derived from all 4 individuals. Additionally, I also see evidence of sites that show a pattern of inheritance and sites that are unique to each individual (Figure 6-11). There are similar numbers of unique H3K4me3 sites in all members of the family, with the exception of the cells derived from the euploid father, which have 10x as many unique sites. This may be due to differences in crosslinking or sheering in this sample. Together these data, with the GABPA ChIP-seq data, suggest that ChIP-seq may not be quantitative enough to see variations in the genome-wide binding profiles of transcriptional regulators in cases of relatively small dosage shift (as would be expected in Down syndrome).

Cell line	Peaks
T21	25042
2N Brother	24366
2N Mom	22100
2N Dad	26608
Total	**31176**

Table 6-2 H3K4me3 peak calling summary.

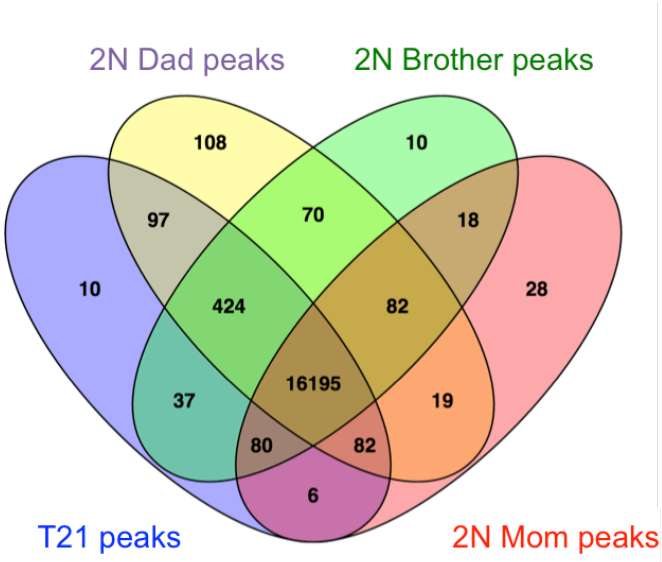


Figure 6-10 The majority of H3K4me3 peaks are shared between individuals. Venn diagram comparing the location of peaks identified in the familial derived lymphoblastoid cell lines. Venn diagram created with Venny (Oliveros 2015).

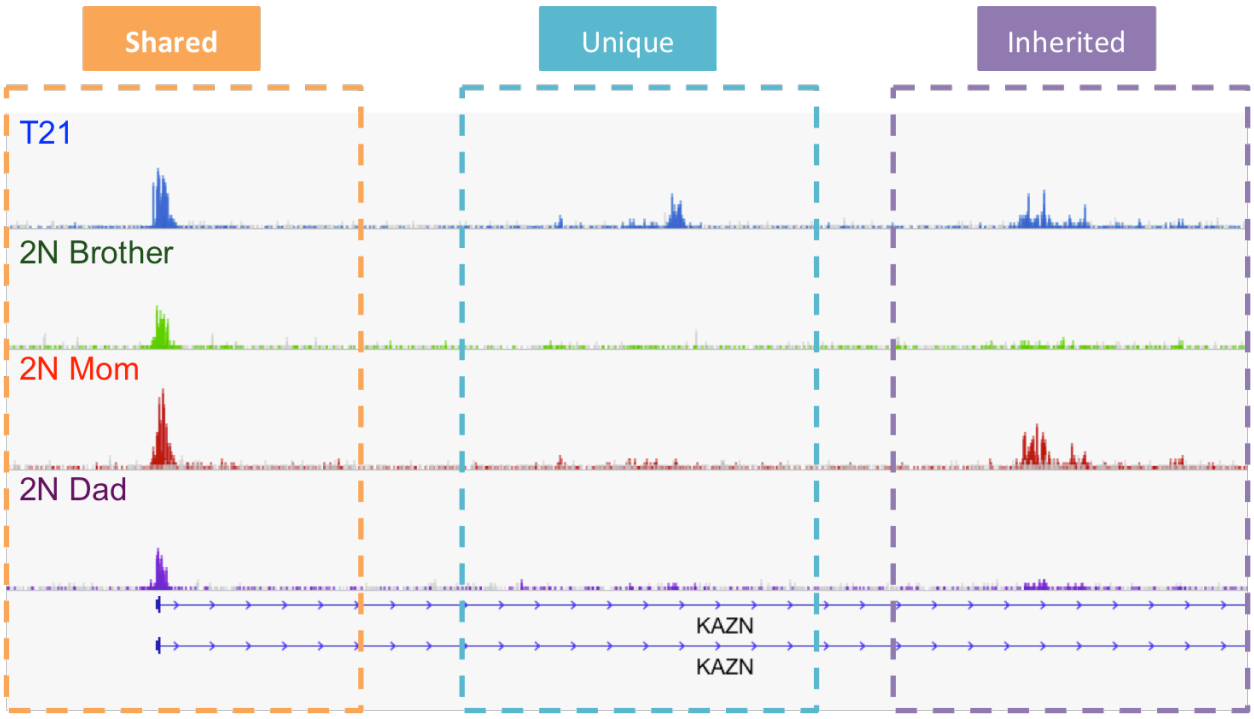


Figure 6-11 Differing patterns of H3K4me3 in the familial derived LCLs. IGV snapshot of H3K4me3 ChIP-seq data for the cells derived from T21 (blue), 2N brother (green), 2N mom (red), and 2N dad (purple) compared to rabbit IgG (grey). Examples of H3K4me3 patterns for shared (orange), unique (cyan), or inherited (light purple) peaks are indicated.

6.2.4 Transcription of binding sites for TFs encoded on HSA21 are increased in T21

Due to the limitations in ChIP-seq, I took a second approach to assess the impact of increased gene dosage from T21 on the function of transcription factors encoded on HSA21. Several studies have shown that enhancer region of the DNA are transcribed, producing eRNAs. Additionally, Dr. Mary Allen and Dr. Joey Azofeifa have recently shown that transcription over TF binding sites is indicative of increased TF activity (Allen et al. 2014). Using nascent transcription data (performed by Dr. Mary Allen) and publically available ChIP-seq data for BACH1, GABPA, RUNX1, and ETS2, I measured changes in transcription over known TF binding sites in the familial derived

LCLs. My hypothesis is that the cells derived from the individual with T21 will exhibit increased transcription over TF binding sites.

To test this hypothesis, I performed differential transcription analysis on GRO-seq read counts over known GABPA binding sites. I used GRO-seq data performed and analyzed by Dr. Mary Allen in the identical familial derived LCLs grown under the same conditions, at the same time, as the cells used in the previously described experiments. For this study, I chose to use publically available GABPA binding data, rather than my own data, because the public data was performed in at least two biological replicates and fully conforms to ENCODE standards (Landt et al. 2012). High confidence GABPA binding sites for GM12878, an immortalized lymphoblastoid cell line, were obtained from ENCODE (GEO:GSE96120) (Dunham et al. 2012). I counted the number of GRO-seq reads (or eRNAs) over each GABPA binding site normalized for the size of the binding site and total reads mapped (for complete details see 6.3.7). I calculated the $\log_2$ transformed eRNA expression between the T21 and 2N brother. As a control, I also compared eRNA expression between two euploid cell lines, the 2N father and the 2N brother. I compared the distribution of differential eRNA expression between the T21 and 2N cells to the distribution between euploid cells and found that the distribution of $\log_2$ transformed GABPA eRNA expression between the T21 cells and euploid cells is significantly increased compared the euploid cells ($p=2.2x10^{-16}$, paired student t-test). This data supports the hypothesis that the dosage imbalance from T21 causes increased GABPA activity in cells derived from an individual with Down syndrome.

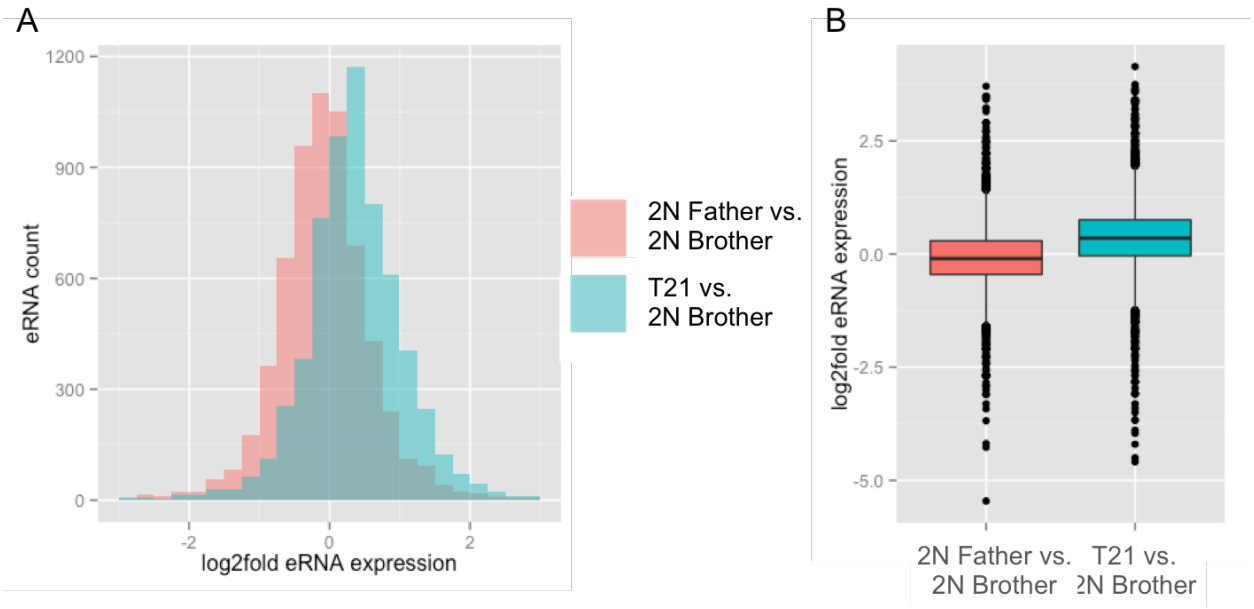


Figure 6-12 Increased GABPA eRNA expression in T21. A) Histogram of the log_2 transformed differential eRNA expression over GABPA binding sites between T21 cells and euploid cells (blue) or euploid cells (red). B) Boxplot of log_2 transformed differential eRNA expression over GABPA binding sites between T21 cells and euploid cells (blue) or euploid cells (red). Transcription over GABPA binding sites is significantly increased in T21 cells compared to euploid ($p=2.2x10^{-16}$)

To see if the increased TF activity was true of other TFs encoded on HSA21, I extended the eRNA expression analysis to BACH1, RUNX1, and ETS2. Each of these TFs are transcribed in LCLs and exhibit a dosage response to T21 (Figure 6-4). Unfortunately, ENCODE binding data is not available for these cells in GM12878 cells (LCLs). Instead I used ChIP-seq for BACH1 (GEO:GSM935576), RUNX1 (GEO:GSE91747), and ETS2 (ENCSR596IKD) that was performed in K562 cells, a cell line derived from a patient with chronic myelogenous leukemia (Koeffler and Golde 1980; Dunham et al. 2012). For each TF tested, I saw an increase in the transcription over the TF binding sites in each sample (Figure 6-13). Although, the eRNA expression is not statistically increased in T21 cells for BACH1, RUNX1, and ETS2, the trend

suggests increased TF activity for HSA21 encoded TFs in T21. The lack of significance in these TFs may be due to using ChIP-seq data from a different cell type than the GRO-seq. While the differential eRNA expression suggests increased TF activity in HSA21 encode TFs, further studies are required to more precisely measure TF activity in T21 and euploid cells.

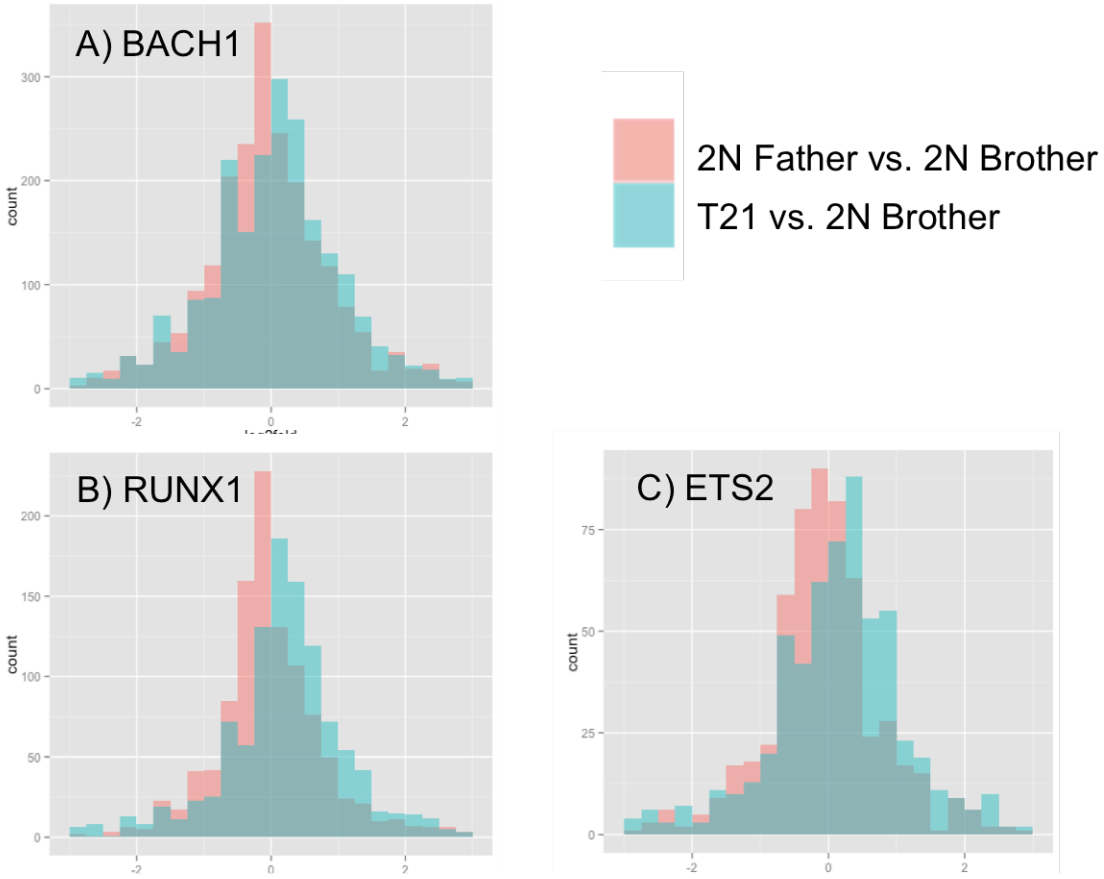


Figure 6-13 Increased HSA21 encoded TF eRNA expression T21 cells. Histogram of the $\log_2$ transformed differential eRNA expression over A) BACH1, B) RUNX1, and C) ETS2 binding sites between T21 cells and euploid cells (blue) or euploid cells (red).

6.3 Methods

6.3.1 Cell growth

Cells were grown according to the encode standards for GM12878, a lymphoblastoid cell line. Cells were grown in T25 tissue coulter flasks with 20mL of media (RPMI 1640 + 2mM L- glutamine + 15% FBS + pen/step) in the upright position at 37°C with 5% CO_2. Cells were passaged ~2-3 days and grown between 0.2 and 1.0 Million cells/mL.

6.3.2 qRT-PCR

Cells were grown to 1.0M cell/ml, washed, and frozen at -80°C. Cell pellets were thawed at RT in TRI reagent (Sigma T9424) and RNA isolated according to manufacture's instructions. 2ug of RNA was reverse transcribed using Multiscribe reverse transcriptase (Thermo Fisher #4311235) with random hexamers accordning to manufactures instructions. cDNA was diluted to 1:100 and quantified using targeted qPCR primers (Table 14-6) and SYBR select (Life Technologies #4472908) on the Biorad CFX qPCR system. A standard curve was used to determine linear range and efficiency of the primers. Gene expression was internally normalized to *RBP13* expression and error propagated for replicates. I tested several genes for internal normalization since we worried that T21 may alter the expression of normalization genes. I got the most reproducibility from RPL13 and 18s, and poor reproducibility from ACTBL2 and GAPDH (data not shown). I chose RPL13 for internal normalization because, unlike 18s, I did not need to de a separate dilution. Primers were designed to span exon-exon junctions to specifically measure mature RNA.

6.3.3 Chromatin Immunoprecipitation (ChIP) library preparation

Chromatin isolation ChIP-seq libraries were prepared using the ENCODE protocol for GM12878 cells from the Myers lab (V011014). Briefly, cells were cross-linked in a final concentration of 1% formaldehyde for 10 minutes at room temperature. To stop the crosslinking reaction, glycine was added to a final concentration of 0.125M. Cells were transferred to 50mL conicals and pelleted for 5min at 5,000xg at 4°C, washed with ice cold 1X PBS and snap frozen in liquid nitrogen in aliquots of $2x10^7$ cells (~1 flask of cells per aliquot). Cells pellets were thawed in 1mL Farnham lysis buffer (5 mM PIPES pH 8.0 / 85 mM KCl / 0.5% NP-40 + Protease inhibitor). Cells were broken open by passage through a 20G needle 20x. Nuclei were collected for 5 min at 2,000xg at 4°C and resuspended in 300uL RIPA (1X PBS / 1% NP-40 / 0.5% sodium deoxycholate / 0.1% SDS + Protease inhibitor). Chromatin was sheered to a median size 300bp in a Diagenode Bioruptor for 3x 10-minute cycles of 30sec ON and 30sec OFF. After each cycle the water was brought back to ice cold by adding ice to the water bath. The amount of sheering necessary should be optimized for each experiment. Chromatin was collected at 14,000 rpm for 15min.

Immunoprecipitation 200ul of magnetic beads (Dynabeads Protein G, Invitrogen 10003D), was washed 2x in 1mL PBS/BSA (1XPBS / 5 mg/ml BSA) and resuspended in 200ul PBS/BSA + 5ul primary antibody. Antibody was coupled to the bead on a rotator for 2hr at 4°C. Antibody coupled beads were washed 3X in PBS/BSA and resuspended in 100ul PBS/BSA. Antibody coupled beads were added to 300ul sheered chromatin and incubated in a rotator overnight at 4°C. Beads were washed 5x with LiCl Wash Buffer (100 mM Tris pH 7.5 / 500 mM LiCl / 1% NP-40 / 1% sodium

deoxycholate) for 3 minutes each followed by 1X wash with TE Buffer (10 mM Tris-HCl pH 7.5 / 0.1 mM Na_2EDTA). Beads were resuspended in 100ul IP Elution buffer (1% SDS / 0.1 M $NaHCO_3$) and chromatin eluted at 65°C for 1 hour. Beads were collected at 14,000rpm at RT for 3min and supernatant transferred to a new tube. Crosslinks were reversed overnight at 65°C. Chromatin was isolated using the QIAquick PCR Purification Kit and eluted in two aliquots of 30ul EB buffer warmed to 55°C. DNA was measured with a Qubit.

Library prep ChIP-seq libraries were prepared from 10ng ChIP'd DNA as input using the NEBNext DNA library prep kit (E6040) following manufacture's instructions. Libraries were multiplexed and sequenced on an Illumina HiSeq 2000 on a 1x50 flow-cell.

6.3.4 ChIP sequencing analysis

Read Mapping ChIP-seq reads were mapped to the hg19 genome. The reads were mapped using the Bowtie2 v2.0.2 (Langmead and Salzberg 2012) local alignment strategy, allowing for multiple mapping, and setting with default options. The mapped reads then underwent file format conversion into the binary format for downstream analysis using Samtools view, sort, and index v0.1.18 (Li et al. 2009). Post-alignment to the genome, duplicate pairs resulting from PCR over-amplification was removed using Samtools rmdup. Read mapping is summarized in Table 6-3.

Antibody	Cell Line	Reads	Unique Mapped	
GABPA	Ethan (T21)	2.48E+07	2.14E+07	87.28%
	Eric (2N Brother)	2.79E+07	2.46E+07	89.83%
	Elizabeth (2N Mom)	2.45E+07	2.12E+07	87.99%
	Eli (2N Dad)	2.38E+07	1.92E+07	82.09%
Mus IgG	Ethan (T21)	2.18E+07	1.86E+07	86.20%
	Eric (2N Brother)	2.42E+07	2.12E+07	88.72%
	Elizabeth (2N Mom)	2.58E+07	2.17E+07	85.34%
	Eli (2N Dad)	1.98E+07	1.65E+07	84.51%
H3K4me3	Ethan (T21)	2.35E+07	2.00E+07	87.52%
	Eric (2N Brother)	1.88E+07	1.64E+07	89.43%
	Elizabeth (2N Mom)	2.16E+07	1.88E+07	89.20%
	Eli (2N Dad)	2.17E+07	1.84E+07	87.23%
Rab IgG	Ethan (T21)	2.24E+07	1.81E+07	82.23%
	Eric (2N Brother)	2.05E+07	1.78E+07	88.81%
	Elizabeth (2N Mom)	1.78E+07	1.39E+07	80.25%
	Eli (2N Dad)	2.32E+07	2.00E+07	88.48%

Table 6-3 ChIP sequencing summary

Peak Calling Peak calling was performed using MACs v2.0.9 (Zhang et al. 2008) with the following settings: -t <ChIP reads> -c <IgG reads> -f BAM –g hs –q 0.01 –n <out folder>. After peak-calling the H3K4me3 files, the H3K4me3 peaks within 1kb were merged to a single peak using bedtools intersectBed v 2.16.2 (Quinlan and Hall 2010).

6.3.5 Minimum peak coverage analysis

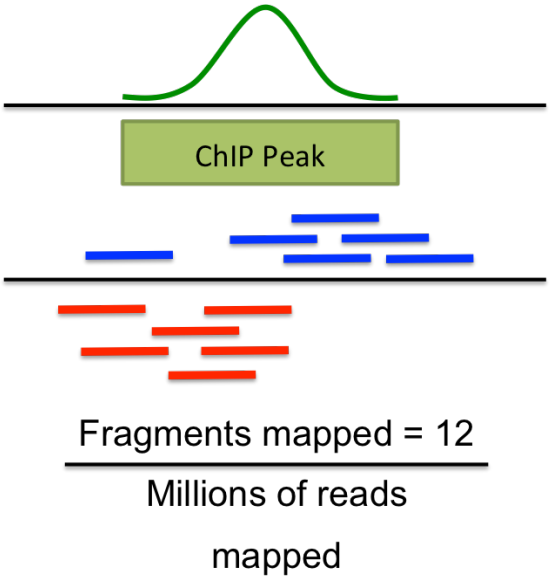
ChIP peaks from each cell line was merged into a single peak file which contained a single entry for each peak found in all samples using bedtools mergeBed v 2.16.2 (Quinlan and Hall 2010). The number of reads over each peak was counted using bedtoold coverageBed v2.16.2 with default settings. Additionally, the number of reads in the control data (IgG) was also counted for each peak. Fragments mapped per kilobase per millions reads mapped (FPKM) was calculated using a custom script. The median FPKM was determined for each dataset using a custom python script and an FPKM of 2.23 was determined to be the threshold for peaks above background.

6.3.6 Differential ChIP occupancy analysis

The number of reads over ChIP region was determined as described in 6.3.5. Differential occupancy analysis was performed using DESeq v1.10.1 (Anders and Huber 2010). Since there were no biological replicates, the “blind” method was used to estimate dispersions. Differential occupancy was determined for all evolved clones against the diploid ancestral strain. Genes were considered differentially expressed between two strains if their adjusted p-value was less than 0.05.

6.3.7 eRNA expression counts

High quality ChIP-seq peak files were obtained from the ENCODE consortium for GABPA, BACH1, RUNX1, and ETS2 (Dunham et al. 2012). The number of GRO-seq reads (sequencing and analysis performed by Dr. Mary Allen) over each TF peak was counted using bedtools coverageBed v2.16.2 with default settings. The normalized read count of FPKM was calculated for each peak using a custom python script, summarized in Figure 6-14.



Fragments mapped = 12

Millions of reads mapped

GABPA	peak4897	FPKM
Eli (Dad)		1.41
Elizabeth (Mom)		1.92
Eric (Brother)		0.39
Ethan (DS)		3.11

Figure 6-14 Diagram for calculating transcription over a ChIP-seq binding site

6.4 Conclusions

I hypothesized than an increase in TF dosage due to an additional copy of chromosome 21 would alter the genome-wide binding profile of TFs encoded on HSA21. I initially tested this hypothesis using ChIP-sequencing for GABPA, a TF encoded on HSA21 and H3K4me3, a mark for active promoters. The ChIP-sequencing data did not show significant alterations in the location or occupancy in the binding profiles of the transcriptional regulators. The binding profiles were most heavily influenced by the variation between samples rather than the biological differences within the samples. Our data suggests that ChIP-seq lacks sufficient resolution, particularly when comparing between separate samples, to measure small alterations in the binding profiles of proteins.

Alternatively, I measured transcription factor activity in the familial derived LCLs through differential expression of eRNA, or enhancer RNAs, over known TF binding sites. Our lab has shown that eRNA expression is a better measure of TF activity than TF binding (unpublished). Overall, I saw an increase in transcription over TF binding sites for TFs encoded on HSA21 in the cells isolated from the individual with T21 compared to euploid cells. However, the increase in transcription was not significant in for all TFs analyzed, suggesting that increased DNA dosage may not have an effect on the activity of all TFs encoded on HSA21. Additionally, lack of significant differences may be an artifact of using ChIP-seq data from a different cell line than the transcriptional data. In the future, these experiments would be improved by an increased number of replicates (only a single replicate of GRO-seq was analyzed) and matched GRO-seq and ChIP-seq data from the same cell lines. Tfit, a ChIP-free method

for determining eRNAs using GRO-seq data (by Dr. Joey Azofeifa), can also be applied to the these studies to determine differential eRNA expression (including novel sites of eRNA expression). This work is currently ongoing by Dr. Mary Allen.

While the differential eRNA expression results are promising, more careful experiments are needed to fully understand the effect of TF dosage on its genome-wide function. For example, one of my major concerns is that all TFs tested show in increase in eRNAs over TF binding sites. At least 3 replicates with synthetic RNA spike-ins should be used to normalize the data before major conclusions can be drawn from the eRNA differential expression data. In particular, I am concerned that fragmentation differences between libraries could affect the number of eRNA (particularly short RNAs) sequenced. In future experiments, I would like to measure differential expression for TFs encoded on non-HSA21 chromosomes as a control.

Another difficulty in these experiments was comparing alterations between cells derived from different individuals rather than isogenic cell lines. Differential ChIP or expression analyses are usually performed in the same cell line before and after treatment. Ideally these studies would be performed in cell lines that are isogenic except for HSA21. While human cells lines do exist for T21 and an isogenic euploid, we have not been able to get access to these lines in our lab (Letourneau et al. 2014). Instead, mouse cells, such as embryonic fibroblasts, could be used in lieu of human cell lines. I believe that using a mouse model offers significant advantages to human cell line studies: isolated mouse cells are primary cells, not immortalized, and there are multiple aneuploid models in which to study gene dosage effects of TF activity.

7 REFERENCES

Adams J, Hansche PE. 1974. Population studies in microorganisms. Genetics 76:327–335.

Adams J, Rosenzweig F. 2014. Experimental microbial evolution: history and conceptual underpinnings. Genomics 104:393–398.

Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. 8:135–141.

Allen MA nn, Andrysik Z, Dengler VL, Mellert HS, Guarnieri A, Freeman JA, Sullivan KD, Galbraith MD, Luo X, Kraus WL, et al. 2014. Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. Elife 3:e02200.

Anders KR, Kudrna JR, Keller KE, Kinghorn B, Miller EM, Pauw D, Peck AT, Shellooe CE, Strong IJT. 2009. A strategy for constructing aneuploid yeast strains by transient nondisjunction of a target chromosome. BMC Genet. 10:36.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106.

Anderson JB, Sirjusingh C, Parsons AB, Boone C, Wickens C, Cowen LE, Kohn LM. 2003. Mode of Selection and Experimental Evolution of Antifungal Drug Resistance in Saccharomyces cerevisiae. Genetics 163:1287–1298.

Anderson JB, Sirjusingh C, Ricker N. 2004. Haploidy, diploidy and evolution of antifungal drug resistance in Saccharomyces cerevisiae. Genetics 168:1915–1923.

Antonarakis SE. 2016. Down syndrome and the complexity of genome dosage imbalance. Nat. Rev. Genet. 18:147–163.

Araya CL, Payen C, Dunham MJ, Fields S. 2010. Whole-genome sequencing of a laboratory- evolved yeast strain.

Atwood KC, Schneider LK, Ryan FJ. 1951. Periodic Selection in Escherichia coli. PNAS 37:146–155.

Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. Nat. Rev. Genet. 14:827–839.

Bell G, Gonzalez A. 2009. Evolutionary rescue can prevent extinction following environmental change. Ecol. Lett. 12:942–948.

Birchler JA, Bhadra U, Bhadra MP, Auger DL. 2001. Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits. Dev. Biol. 234:275–88.

Bjedov I, Tenaillon O, Gérard B, Souza V, Denamur E, Radman M, Taddei F, Matic I. 2003. Stress-induced mutagenesis in bacteria. Science 300:1404–9.

Bleuven C, Landry CR. 2016. Molecular and cellular bases of adaptation to a changing environment in microorganisms. Proc. R. Soc. B Biol. Sci. 283:20161458.

Bloom SE. 1972. in Chicken (Gallus domesticus) Embryos : Types , Frequencies and Phenotypic Effects. Chromosoma 37:309–326.

Boer VM, Winde JH De, Pronk JT, Piper MDW. 2003. The Genome-wide Transcriptional Responses of Saccharomyces cerevisiae Grown on Glucose in Aerobic Chemostat Cultures Limited for Carbon , Nitrogen , Phosphorus , or Sulfur. J. Biol. Chem. 278:3265–3274.

Boles E, Hollenberg CP. 1997. The molecular genetics of hexose transport in yeasts. FEMS Microbiol. Rev. 21:85–111.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bonhoure N, Bounova G, Bernasconi D, Praz V, Lammers F, Canella D, Willis IM, Herr W, Hernandez N, Delorenzi M, et al. 2014. Quantifying ChIP-seq data: A spiking method providing an internal reference for sample-to-sample normalization. Genome Res. 24:1157–1168.

Brown CJ, Todd KM, Rosenzweig RF. 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol. Biol. Evol. 15:931–942.

Carlson M, Botstein D. 1982. Two differentially regulated mRNAs with different 5′ ends encode secreted and intracellular forms of yeast invertase. Cell 28:145–154.

Chen G, Bradford WD, Seidel CW, Li R. 2012. Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. Nature 482:246–250.

Chen G, Mulla WA, Kucharavy A, Tsai HJ, Rubinstein B, Conkright J, McCroskey S, Bradford WD, Weems L, Haug JS, et al. 2015. Targeting the adaptability of heterogeneous aneuploids. Cell 160:771–784.

Chen ZJ, Ni Z. 2006. Mechanisms of genomic rearrangements and gene expression changes in plant polyploids. BioEssays 28:240–252.

Clemons K V, Park P, McCusker JH, McCullough MJ, Davis RW, Stevens DA. 1997. Application of DNA typing methods and genetic analysis to epidemiology and taxonomy of Saccharomyces isolates. J Clin Microbiol 35:1822–1828.

Colvin KL, Yeager ME. 2017. What people with Down Syndrome can teach us about cardiopulmonary disease. Eur. Respir. Rev. 26:160098.

Comai L. 2005. The advantages and disadvantages of being polyploid. Nat. Rev. Genet. 6:836–846.

Cook M, Tyers M. 2007. Size control goes global. Curr. Opin. Biotechnol. 18:341–350.

Cooper TF, Rozen DE, Lenski RE. 2003. Parallel changes in gene expression after 20 , 000 generations of evolution in Escherichia coli. PNAS 100:1072–1077.

Costa V, Angelini C, D'Apice L, Mutarelli M, Casamassimi A, Sommese L, Gallo MA, Aprile M, Esposito R, Leone L, et al. 2011. Massive-scale RNA-Seq analysis of non ribosomal transcriptome in human trisomy 21. PLoS One 6:e18493.

Cowen LE, Sanglard D, Calabrese D, Sirjusingh C, Anderson JB, Kohn LM. 2000. Evolution of Drug Resistance in Experimental Populations of Candida albicans. Jounral Bacteriol. 182:1515–1522.

Creasy MK, Crolla JA, Alberman ED. 1976. A Cytogenetic Study of Human Spontaneous Abortions Using Banding Techniques. Hum. Genet. 31:177–196.

Cross FR. 1997. "Marker Swap" Plasmids : Convenient Tools for Budding Yeast Molecular Genetics. Yeast 13:647–653.

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, et al. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res. 16:738–749.

Dalca A V., Rumble SM, Levy S, Brudno M. 2010. VARiD: A variation detection framework for color-space and letter-space platforms. Bioinformatics 26:343–349.

Darwin C. 1859. On the Origin of Species by Means of Natural Selection. London: John Murry.

DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498.

Dickson RC, Nagiec EE, Wells GB, Nagiec MM, Lester RL. 1997. Synthesis of Mannose-(inositol-P) 2 -ceramide, the Major Sphingolipid in Saccharomyces cerevisiaw, Resquires the IPT1 (YDR072c) Gene. Biochemistry 272:29620–29625.

Dierssen M. 2012. Down syndrome: the brain in trisomic mode. Nat. Rev. Neurosci. 13:844–58.

Dietzel KL, Ramakrishnan V, Murphy EE, Bisson LF. 2012. MTH1 and RGT1 demonstrate combined haploinsufficiency in regulation of the hexose transporter genes in Saccharomyces cerevisiae. BMC Genet. 13:107.

Dobzhansky T, Pavlovsky O. 1957. An Experimental Study of Interaction between Genetic Drift and Natural Selection. Evolution (N. Y). 11:311–319.

Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, et al. 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74.

Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F, Botstein D. 2002. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. PNAS 99:16144–16149.

Elena SF, Lenski RE. 2003. Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. 4:457–469.

Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, Dwight SS, Hitz BC, Karra K, Nash RS, et al. 2014. The Reference Genome Sequence of Saccharomyces cerevisiae : Then and Now. G3 Genes, Genomes, Genet. Genes|Genomes|Genetics 4:389–398.

Ezov KT, Boger-Nadjar E, Frenkel Z, Katsperovski I, Kemeny S, Nevo E, Korol A, Kashi Y. 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.

Ferea TL, Botstein D, Brown PO, Rosenzweig RF. 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. PNAS 96:9721–6.

Fisher KJ, Lang GI. 2016. Experimental evolution in fungi: An untapped resource. Fungal Genet. Biol. 94:88–94.

Fitzpatrick DR. 2005. Transcriptional consequences of autosomal trisomy : primary gene dosage with complex downstream effects. 21.

Flick KM, Spielewoy N, Kalashnikova TI, Guaderrama M, Zhu Q, Chang H-C, Wittenberg C. 2003. Grr1-dependent Inactiavtion of MTH1 Mediates Glucose-induced Dissociation of Rgt1 from HTX Gene Promoters. Mol. Biol. Cell 14:3230–3241.

Fonatsch C. 2010. The role of chromosome 21 in hematology and oncology. Genes, Chromosom. Cancer 49:497–508.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531–1545.

Forche A, Abbey D, Pisithkul T, Weinzierl MA, Ringstrom T, Bruck D, Petersen K, Berman J. 2011. Stress alters rates and types of loss of heterozygosity in candida albicans. MBio 2:1–9.

Fujiwara T, Bandi M, Nitta M, Ivanova E V, Bronson RT, Pellman D. 2005. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. Nature 437:1043–7.

Futuyama D. 1970. Variation in Genetic Response to Interspecific Competition in Laboratory Populations of Drosophila. Am. Nat. 104:239–252.

Galhardo RS, Hastings PJ, Rosenberg SM. 2007. Mutation as a Stress Response and the Regulation of Evolvability. Crit. Rev. Biochem. Mol. Biol. 42:399–435.

Galitski T, Saldanha a J, Styles C a, Lander ES, Fink GR. 1999. Ploidy regulation of gene expression. Science 285:251–254.

Gancedo JM. 2008. The early steps of glucose signalling in yeast. FEMS Microbiol. Rev. 32:673–704.

Ganem NJ, Storchova Z, Pellman D. 2007. Tetraploidy, aneuploidy and cancer. Curr. Opin. Genet. Dev. 17:157–162.

Gardiner K. 2006. Transcriptional dysregulation in Down syndrome: predictions for altered protein complex stoichiometries and post-translational modifications, and consequences for learning/behavior genes ELK, CREB, and the estrogen and glucocorticoid receptors. Behav. Genet. 36:439–53.

Garland T, Adolph SC. 1994. Why Not to Do Two-Species Comparative Studies : Limitations on Inferring Adaptation. Physiol. Zool. 67:797–828.

Gasch a P, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11:4241–4257.

Gasch AP. 2002. The environmental stress response : a common yeast response to diverse environmental stresses. Top. Curr. Genet. 1:11–70.

Gasch AP, Werner-Washburne M. 2002. The genomics of yeast responses to environmental stress and starvation. Funct. Integr. Genomics 2:181–192.

Gazave E, Darre F, Morcillo-Suarez C, Petit-Marty N, Carreno A, Marigorta UM, Ryder O a, Blancher A, Rocchi M, Bosch E, et al. 2011. Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. Genome Res. 21:1626–1639.

Gerstein AC, Chun HJE, Grant A, Otto SP. 2006. Genomic convergence toward diploidy in Saccharomyces cerevisiae. PLoS Genet. 2:1396–1401.

Gerstein AC, Cleathero LA, Mandegar MA, Otto SP. 2011. Haploids adapt faster than diploids across a range of environments. J. Evol. Biol. 24:531–540.

Gerstein AC, Lim H, Berman J, Hickman MA. 2017 Feb. Ploidy tug-of-war: evolutionary and genetic environments influence the rate of ploidy drive in a human fungal pathogen. Evolution (N. Y).:1–27.

Gerstein AC, Mcbride RM, Otto SP. 2008. Ploidy reduction in Saccharomyces cerevisiae. :91–94.

Gerstein AC, Otto SP. 2009. Ploidy and the causes of genomic evolution. J. Hered. 100:571–81.

de Godoy LMF, Olsen J V, Cox J, Nielsen ML, Hubner NC, Fröhlich F, Walther TC, Mann M. 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455:1251–1254.

Goldstein a L, McCusker JH. 1999. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast 15:1541–53.

Gorla GR, Malhi H, Gupta S. 2001. Polyploidy associated with oxidative injury attenuates proliferative potential of cells. J. Cell Sci. 114:2943–2951.

Gresham D, Desai MM, Tucker CM, Jenq HT, Pai D a., Ward A, DeSevo CG, Botstein D, Dunham MJ. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 4.

Grishin A V., Rothenberg M, Downs MA, Blumer KJ. 1998. Mot3, a Zn finger transcription factor that modulates gene expression and attenuates mating pheromone signaling in Saccharomyces cerevisiae. Genetics 149:879–892.

Halfmann R, Jarosz DF, Jones SK, Chang A, Lancaster AK, Lindquist S. 2012. Prions are a common mechanism for phenotypic inheritance in wild yeasts. Nature 482:363–8.

Hallstrom TC, Lambert L, Schorling S, Balzi E, Goffeau A, Moye-Rowley WS. 2001. Coordinate Control of Sphingolipid Biosynthesis and Multidrug Resistance in Saccharomyces cerevisiae. J. Biol. Chem. 276:23674–23680.

Harrison BD, Hashemi J, Bibi M, Pulver R, Bavli D, Nahmias Y, Wellington M, Sapiro G, Berman J. 2014. A Tetraploid Intermediate Precedes Aneuploid Formation in Yeasts Exposed to Fluconazole. PLoS Biol. 12.

Hartley D, Blumenthal T, Carrillo M, DiPaolo G, Esralew L, Gardiner K, Granholm AC, Iqbal K, Krams M, Lemere C, et al. 2015. Down syndrome and Alzheimer's disease: Common pathways, common goals. Alzheimer's Dement. 11:700–709.

Harvey PH, Purvis A. 1991. Comparative Methods for Explaining Adaptations. Nature 351:619–624.

Hass JWJ. 2000. THE REVEREND DR WILLIAM HENRY DALLINGER, F.R.S. (1839–1909) by J.W. H. Notes Rec. R. Soc. Lond. 54:53–65.

He X, Zhang J. 2005. Gene complexity and gene duplicability. Curr. Biol. 15:1016–1021.

Hegreness M. 2006. An Equivalence Principle for the Incorporation of Favorable Mutations in Asexual Populations. Science 311:1615–1617.

Hittinger CT, Gonçalves P, Sampaio JP, Dover J, Johnston M, Rokas A. 2010. Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature 464:54–58.

Hixon ML, Obejero-Paz C, Muro-Cacho C, Wagner MW, Millie E, Nagy J, Hassold TJ, Gualberto A. 2000. Cks1 mediates vascular smooth muscle cell polyploidization. J. Biol. Chem. 275:40434–40442.

Holmes DL, Lancaster AK, Lindquist S, Halfmann R. 2013. Heritable Remodeling of Yeast Multicellularity by an Environmentally Responsive Prion. Cell 153:153–165.

Homer N, Merriman B, Nelson SF. 2009. BFAST: An alignment tool for large scale genome resequencing. PLoS One 4:e7767.

Homer N, Nelson SF. 2010. Improved variant discovery through local re-alignment of short-read next-generation sequencing data using SRMA. Genome Biol. 11:R99.

Hong J, Gresham D. 2014. Molecular Specificity, Convergence and Constraint Shape Adaptive Evolution in Nutrient-Poor Environments. Fay JC, editor. PLoS Genet. 10:e1004041.

Hongay C, Jia N, Bard M, Winston F. 2002. Mot3 is a Transcriptional Repressor of Ergosterol Biosynthesis Genes and is Required for Normal Vacuolar Function in Saccharomyces cerevisiae. Embo J. 21:4114–4124.

Iskow RC, Gokcumen O, Lee C. 2012. Exploring the role of copy number variants in human adaptation. Trends Genet. 28:245–257.

Johnston M. 1999. Feasting , fasting and fermenting. TIG 15:29–33.

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

Kim J, Polish J, Johnston M. 2003. Specificity and Regulation of DNA Binding by the Yeast Glucose Transporter Gene Respressor Rgt1. Mol. Cell. Biol. 23:5208–5216.

King KC, Seppala O, Neiman M. 2012. Is more better? Polyploidy and parasite resistance. Biol. Lett. 8:598–600.

Koeffler HP, Golde DW. 1980. Human Myeloid Leukemia Cell Lines: A Review. Blood 56:344–350.

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

Koschwanez JH, Foster KR, Murray AW. 2013. Improved use of a public good selects for the evolution of undifferentiated multicellularity. Elife 2013:1–27.

Koszul R, Caburet S, Dujon B, Fischer G. 2004. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. EMBO J. 23:234–43.

Kouzarides T. 2007. Chromatin modifications and their function. Cell 128:693–705.

Krogerus K, Arvas M, De Chiara M, Magalhães F, Mattinen L, Oja M, Vidgren V, Yue JX, Liti G, Gibson B. 2016. Ploidy influences the functional attributes of de novo lager yeast hybrids. Appl. Microbiol. Biotechnol. 100:7203–7222.

Krogh a, Larsson B, von Heijne G, Sonnhammer E. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305:567–580.

Kruckeberg a L, Bisson LF. 1990. The HXT2 gene of Saccharomyces cerevisiae is required for high-affinity glucose transport. Mol. Cell. Biol. 10:5903–5913.

Kvitek DJ, Sherlock G. 2011. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet. 7:e1002056.

Kvitek DJ, Sherlock G. 2013. Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. PLoS Genet. 9:e1003972.

Lafuente MJ, Gancedo C, Jauniaux JC, Gancedo JM. 2000. Mth1 receives the signal given by the glucose sensors Snf3 and Rgt2 in Saccharomyces cerevisiae. Mol. Microbiol. 35:161–172.

Lagunas R. 1993. Sugar transport in Saccharomyces cerevisiae. FEMS Microbiol. Lett. 104:229–242.

Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet J, Subramanian A, Ross KN, et al. 2006. The Connectivity Map : Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. Science 313:1929–1935.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, et al. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 22:1813–31.

Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, Desai MM. 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500:571–4.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Lee HO, Davidson JM, Duronio RJ. 2010. Endoreplication : polyploidy with purpose. Genes Dev. 23:2461–2477.

Letourneau A, Santoni F a, Bonilla X, Sailani MR, Gonzalez D, Kind J, Chevalier C, Thurman R, Sandstrom RS, Hibaoui Y, et al. 2014. Domains of genome-wide gene expression dysregulation in Down's syndrome. Nature 508:345–50.

Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods 7:709–15.

Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature 519:181–186.

Li H, Cherry S, Klinedinst D, DeLeon V, Redig J, Reshey B, Chin MT, Sherman SL, Maslen CL, Reeves RH. 2012. Genetic modifiers predisposing to congenital heart disease in the sensitized down syndrome population. Circ. Cardiovasc. Genet. 5:301–308.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Liang H, Gaber RF. 1996. A novel signal transduction pathway in Saccharomyces cerevisiae defined by Snf3-regulated expression of HXT6. Mol. Biol. Cell 7:1953–1966.

Liang H, Gaber RF. 1996. A novel signal transduction pathway in Saccharomyces cerevisiae defined by Snf3-regulated expression of HXT6. Mol. Biol. Cell 7:1953–1966.

Lin Z, Li WH. 2011. Expansion of hexose transporter genes was associated with the evolution of aerobic fermentation in yeasts. Mol. Biol. Evol. 28:131–142.

Liu H, Stylest CA, Fink GR, Rudolph AA. 1996. Saccharomyces cerevisiae S288C Has a Mutation in FL08, a Gene Required for Filamentous Growth. Genetics 144:967–978.

Liu W, Zhou H, Liu L, Zhao C, Deng Y, Chen L, Wu L, Mandrycky N, McNabb CT, Peng Y, et al. 2015. Disruption of neurogenesis and cortical development in transgenic mice misexpressing Olig2, a gene in the Down syndrome critical region. Neurobiol. Dis. 77:106–116.

Longtine MS, Mckenzie III A, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast 14:953–961.

Lott IT, Dierssen M. 2010. Cognitive deficits and associated neurological complications in individuals with Down's syndrome. Lancet Neurol. 9:623–633.

Lu Y-J, Swamy KBS, Leu J-Y. 2016. Experimental Evolution Reveals Interplay between Sch9 and Polyploid Stability in Yeast. PLOS Genet. 12:e1006409.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151–5.

Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473.

Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. PNAS 105:9272–9277.

Mable BK. 2001. Ploidy evolution in the yeast Saccharomyces cerevisiae: A test of the nutrient limitation hypothesis. J. Evol. Biol. 14:157–170.

Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y. 2005. Modeling gene and genome duplications in eukaryotes. PNAS 102:5454–9.

Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. 2012. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science 337:1190–1195.

Mayer VW, Aguilera a. 1990. High levels of chromosome instability in polyploids of Saccharomyces cerevisiae. Mutat. Res. 231:177–186.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Megarbane A, Ravel A, Mircher C, Sturtz F, Grattau Y, Rethore MO, Delabar J-M, Mobley WC. 2009. The 50th anniversary of the discovery of trisomy 21: the past, present, and future of research and treatment of Down syndrome. Genet. Med. 11:611–616.

Mishra S, Whetstine JR. 2016. Different Facets of Copy Number Changes: Permanent, Transient, and Adaptive. Mol. Cell. Biol. 36:1050–63.

Møller HD, Parsons L, Jørgensen TS, Botstein D, Regenberg B. 2015. Extrachromosomal circular DNA is common in yeast. PNAS 112:E3114-22.

Moriya H, Johnston M. 2004. Glucose sensing and signaling in Saccharomyces cerevisiae through the Rgt2 glucose sensor and casein kinase I. PNAS 101:1572–1577.

Mulla W, Zhu J, Li R. 2014. Yeast: a simple model system to study complex phenomena of aneuploidy. FEMS Microbiol. Rev. 38:201–212.

Muller LAH, McCusker JH. 2009. Microsatellite analysis of genetic diversity among clinical and nonclinical Saccharomyces cerevisiae isolates suggests heterozygote advantage in clinical environments. Mol. Ecol. 18:2779–2786.

Mumberg D, Müller R, Funk M. 1995. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene 156:119–122.

Nagaoka SI, Hassold TJ, Hunt PA. 2012. Human aneuploidy: mechanisms and new insights into an age-old problem. Nat Rev Genet 13:493–504.

Naumov G, Turakainen H, Naumova E, Aho S, Korhola M. 1990. A new family of polymorphic genes in Saccharomyces cerevisiae: α-galactosidase genes MEL1-MEL7. Mol. Gen. Genet. 224:119–128.

New AM, Cerulus B, Govers SK, Perez-Samper G, Zhu B, Boogmans S, Xavier JB, Verstrepen KJ. 2014. Different Levels of Catabolite Repression Optimize Growth in Stable and Variable Environments. PLoS Biol. 12:17–20.

Nižetić D, Groet J. 2012. Tumorigenesis in Down's syndrome: big lessons from a small chromosome. Nat. Rev. Cancer 12:721–732.

Novocraft. Novocraft short read alignment package. 2009.

Oberringer M, Lothschütz D, Jennewein M, Koschnick M, Mutschler W, Hanselmann RG. 1999. Centrosome multiplication accompanies a transient clustering of polyploid cells during tissue repair. Mol. Cell Biol. Res. Commun. 2:190–196.

Ohno S. 1970. Evolution by Gene Duplication. Berlin, Heidelberg: Springer Berlin Heidelberg.

Oliveros JC. 2015. Venny. An interactive tool for comparing lists with Venn's diagrams.

Orlando D a, Chen MW, Bradner JE, Guenther MG, Orlando D a, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER. 2014. Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome Resource Quantitative ChIP-Seq Normalization Reveals Global Modulation of the Epigenome. CellReports 9:1163–1170.

Oromendia AB, Amon A. 2014. Aneuploidy: implications for protein homeostasis and disease. Dis. Model. Mech. 7:15–20.

Oromendia AB, Dodgson SE, Amon A. 2012. Aneuploidy causes proteotoxic stress in yeast. Genes Dev. 26:2696–708.

Orr HA. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Orr HA, Otto SP. 1994. Does diploidy increase the rate of adaptation? Genetics 136:1475–1480.

Osborn TC, Chris Pires J, Birchler JA, Auger DL, Chen ZJ, Lee HS, Comai L, Madlung A, Doerge RW, Colot V, et al. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends Genet. 19:141–147.

Otto SP. 2007. The Evolutionary Consequences of Polyploidy. Cell 131:452–462.

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

Ozcan S, Dover J, Rosenwald AG, Wölfl S, Johnston M. 1996. Two glucose transporters in Saccharomyces cerevisiae are glucose sensors that generate a signal for induction of gene expression. PNAS 93:12428–32.

Ozcan S, Johnston M. 1995. Three different regulatory mechanisms enable yeast hexose transporter (HXT) genes to be induced by different levels of glucose. Mol. Cell. Biol. 15:1564–72.

Ozcan S, Johnston M. 1999. Function and regulation of yeast hexose transporters. Microbiol. Mol. Biol. Rev. 63:554–569.

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

Pasula S, Chakraborty S, Choi JH, Kim J-H. 2010. Role of casein kinase 1 in the glucose sensor-mediated signaling pathway in yeast. BMC Cell Biol. 11:17.

Pasula S, Jouandot D, Kim JH. 2007. Biochemical evidence for glucose-independent induction of HXT expression in Saccharomyces cerevisiae. FEBS Lett. 581:3230–3234.

Pavelka N, Rancati G, Li R. 2010. Dr Jekyll and Mr Hyde: role of aneuploidy in cellular adaptation and cancer. Curr. Opin. Cell Biol. 22:809–15.

Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R. 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature 468:321–325.

Payen C, Sunshine AB, Ong GT, Pogachar JL, Zhao W, Dunham MJ. 2016. High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations. PLOS Genet. 12:e1006339.

Van de Peer Y. 2004. Computational approaches to unveiling ancient genome duplications. Nat. Rev. Genet. 5:752–63.

Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. Nat. Rev. Genet. 10:725–732.

Polish J a., Kim J-H, Johnston M. 2005. How the Rgt1 Transcription Factor of Saccharomyces cerevisiae Is Regulated by Glucose. Genetics 169:583–594.

Potter H. 2016. Beyond Trisomy 21: Phenotypic Variability in People with Down Syndrome Explained by Further Chromosome Mis-segregation and Mosaic Aneuploidy. J. Down Syndr. Chromosom. Abnorm. 2:2–5.

Prandini P, Deutsch S, Lyle R, Gagnebin M, Delucinge Vivier C, Delorenzi M, Gehrig C, Descombes P, Sherman S, Dagna Bricarelli F, et al. 2007. Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance. Am. J. Hum. Genet. 81:252–63.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formations in flowering plants. Annu. Rev. Ecol. Syst. 29:467–501.

Rancati G, Pavelka N, Fleharty B, Noll A, Trimble R, Walton K, Perera A, Staehling-Hampton K, Seidel CW, Li R. 2008. Aneuploidy Underlies Rapid Adaptive

Evolution of Yeast Cells Deprived of a Conserved Cytokinesis Motor. Cell 135:879–893.

Raser JM, Shea EKO. 2006. Control of Stochasticity in Eukaryotic Gene Expression Jonathan. Science 304:1811–1814.

Reifenberger E, Boles E, Ciriacy M. 1997. Kinetic characterization of individual hexose transporters of Saccharomyces cerevisiae and their relation to the triggering mechanisms of glucose repression. Eur. J. Biochem 245:324–333.

Ridley M. 1983. The explanation of organic diversity: the comparative method and adaptations for mating. London and New York: Clarendon Press (Oxford University Press).

Rolland F, Winderickx J, Thevelein JM. 2002. Glucose-sensing and -signalling mechanisms in yeast. FEMS Yeast Res. 2:183–201.

Roop JI, Chang KC, Brem RB. 2016. Polygenic evolution of a sugar specialization trade-off in yeast. Nature 530:1–14.

Rosmarin a. 2004. GA-binding protein transcription factor: a review of GABP as an integrator of intracellular signaling and protein–protein interactions. Blood Cells, Mol. Dis. 32:143–154.

Roy A, Jouandot D, Cho KH, Kim JH. 2014. Understanding the mechanism of glucose-induced relief of Rgt1-mediated repression in yeast. FEBS Open Bio 4:105–111.

Roy A, Shin YJ, Cho KH, Kim J-H. 2013. Mth1 regulates the interaction between the Rgt1 repressor and the Ssn6-Tup1 corepressor complex by modulating PKA-dependent phosphorylation of Rgt1. Mol. Biol. Cell 24:1493–503.

Ryan FJ. 1953. Evolution Observed. Sci. Am.:78–83.

Sabina J, Johnston M. 2009. Asymmetric signal transduction through paralogs that comprise a genetic switch for sugar sensing in Saccharomyces cerevisiae. J. Biol. Chem. 284:29635–43.

Saldanha AJ. 2004. Java Treeview - Extensible visualization of microarray data. Bioinformatics 20:3246–3248.

Schoustra SE, Debets AJM, Slakhorst M, Hoekstra RF. 2007. Mitotic Recombination Accelerates Adaptation in the Fungus Aspergillus nidulans. PLoS Genet. 3:e68.

Schulte F, Wieczorke R, Hollenberg CP, Boles E. 2000. The HTR1 gene is a dominant negative mutant allele of MTH1 and blocks Snf3- and Rgt2-dependent glucose signaling in yeast. J. Bacteriol. 182:540–542.

Sellis D, Kvitek DJ, Dunn B, Sherlock G, Petrov DA. 2016. Heterozygote advantage is a common outcome of adaptation in Saccharomyces cerevisiae. Genetics 203:1401–1413.

Selmecki A, Bergmann S, Berman J. 2005. Comparative genome hybridization reveals widespread aneuploidy in Candida albicans laboratory strains. Mol. Microbiol. 55:1553–1565.

Selmecki A, Forche A, Berman J. 2006. Aneuploidy and isochromosome formation in drug-resistant Candida albicans. Science 313:367–70.

Selmecki A, Forche A, Berman J. 2010. Genomic plasticity of the human fungal pathogen Candida albicans. Eukaryot. Cell 9:991–1008.

Selmecki AM, Maruvka YE, Richmond P a, Guillet M, Shoresh N, Sorenson AL, De S, Kishony R, Michor F, Dowell R, et al. 2015. Polyploidy can drive rapid adaptation in yeast. Nature 519:349–352.

Sementchenko VI, Watson DK. 2000. Ets target genes: past, present and future. Oncogene 19:6533–6548.

Sémon M, Wolfe KH. 2007a. Consequences of genome duplication. Curr. Opin. Genet. Dev. 17:505–512.

Sémon M, Wolfe KH. 2007b. Rearrangement rate following the whole-genome duplication in teleosts. Mol. Biol. Evol. 24:860–867.

Sheltzer JM, Torres EM, Dunham MJ, Amon A. 2012. Transcriptional consequences of aneuploidy. 98195.

Shim KS, Ferrando-Miguel R LG. 2003. Aberrant protein expression of transcription factors BACH1 and ERG, both encoded on chromosome 21, in brains of patients with Down syndrome and Alzheimer's disease. J Neural Transm Suppl 67:39–49.

Shor E, Fox CA, Broach JR. 2013. The Yeast Environmental Stress Response Regulates Mutagenesis Induced by Proteotoxic Stress. Jinks-Robertson S, editor. PLoS Genet. 9:e1003680.

Soltis DE, Visger CJ, Soltis PS. 2014. The polyploidy revolution then...and now: Stebbins revisited. Am. J. Bot. 101:1057–1078.

Stebbins GL. 1940. The Significance of Polyploidy in Plant Evolution. Am. Nat. 74:54–66.

Storchová Z, Breneman A, Cande J, Dunn J, Burbank K, O'toole E, Pellman D. 2006. Genome-wide genetic analysis of polyploidy in yeast. Nature 443:541–547.

Storchova Z, Pellman D. 2004. From polyploidy to aneuploidy, genome instability and cancer. Nat. Rev. Mol. Cell Biol. 5:45–54.

Storici F, Lewis LK, Resnick M a. 2001. In vivo site-directed mutagenesis using oligonucleotides. Nat. Biotechnol. 19:773–6.

Storici F, Resnick MA. 2006. The Delitto Perfetto Approach to In Vivo Site-Directed Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast. Methods Enzymol. 409:329–345.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette M a, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS 102:15545–15550.

Sunshine AB, Payen C, Ong GT, Liachko I, Tan KM, Dunham MJ. 2015. The Fitness Consequences of Aneuploidy Are Driven by Condition-Dependent Gene Effects. Barton NH, editor. PLOS Biol. 13:e1002155.

Tang Y-C, Amon A. 2013. Gene copy-number alterations: a cost-benefit analysis. Cell 152:394–405.

The ENCODE Project Consortium. 2011. A User's Guide to the Encyclopedia of DNA Elements (ENCODE). Becker PB, editor. PLoS Biol. 9:e1001046.

Theodoris G, Fong NM, Coons DM, Bisson LF. 1994. High-Copy Suppression of Glucose Transport Defects by HXT4 and Regulatory Elements in the Promoters of the HXT Genes in Saccharomyces cerevisiae. Genetics 137:957–966.

Thompson DA, Desai MM, Murray AW. 2006. Ploidy Controls the Success of Mutators and Nature of Mutations during Budding Yeast Evolution. Curr. Biol. 16:1581–1590.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. Brief. Bioinform. 14:178–192.

Torres EM, Dephoure N, Panneerselvam A, Tucker CM, Whittaker C a, Gygi SP, Dunham MJ, Amon A. 2010. Identification of aneuploidy-tolerating mutations. Cell 143:71–83.

Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A. 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science 317:916–24.

Torres EM, Williams BR, Amon A. 2008. Aneuploidy: cells losing their balance. Genetics 179:737–46.

Toussaint M, Conconi A. 2006. High-throughput and sensitive assay to measure yeast cell growth: a bench protocol for testing genotoxic agents. Nat. Protoc. 1:1922–1928.

Travisano M, Mongold JA, Bennett AF, Lenski RE. 1995. Experimental Tests of the Roles of Adaptation, Chance, and History. Science 267:87–90.

Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, Geiler-Samerotte K, Hérissant L, Blundell JR, Levy SF, et al. 2016. Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. Cell 166:1585–1596.

Vilardell M, Rasche A, Thormann A, Maschke-Dutz E, Pérez-Jurado L a, Lehrach H, Herwig R. 2011. Meta-analysis of heterogeneous Down Syndrome data reveals

consistent genome-wide dosage effects related to neurological processes. BMC Genomics 12:229.

Voordeckers K, Kominek J, Das A, Espinosa-Cantú A, De Maeyer D, Arslan A, Van Pee M, van der Zande E, Meert W, Yang Y, et al. 2015. Adaptation to High Ethanol Reveals Complex Evolutionary Pathways. Zhang J, editor. PLOS Genet. 11:e1005635.

Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, et al. 2006. Genomewide nonadditive gene regulation in arabidopsis allotetraploids. Genetics 172:507–517.

Warnatz H-J, Schmidt D, Manke T, Piccini I, Sultan M, Borodina T, Balzereit D, Wruck W, Soldatov A, Vingron M, et al. 2011. The BTB and CNC homology 1 (BACH1) target genes are involved in the oxidative stress response and in control of the cell cycle. J. Biol. Chem. 286:23521–32.

Weaver BA, Cleveland DW. 2006. Does aneuploidy cause cancer? Curr. Opin. Cell Biol. 18:658–667.

Weiss R, Kukora J, Adams J. 1975. The Relationship Between Enzyme Activity , Cell Geometry, and Fitness. PNAS 72:794–798.

Wendel JF. 2000. Genome evolution in polyploids. Plant Mol. Biol. 42:225–249.

Wenger JW, Piotrowski J, Nagarajan S, Chiotti K, Sherlock G, Rosenzweig F. 2011. Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. PLoS Genet. 7:e1002202.

Wichman H a. 1999. Different Trajectories of Parallel Evolution During Viral Adaptation. Science 285:422–424.

Wijshake T, Malureanu LA, Baker DJ, Jeganathan KB, van de Sluis B, van Deursen JM. 2012. Reduced Life- and Healthspan in Mice Carrying a Mono-Allelic BubR1 MVA Mutation. PLoS Genet. 8.

Wolfe KH. 2001. Yesterday's polyploids and the mystery of diploidization. Nat. Rev. Genet. 2:333–341.

Wolvetang E., Bradfield O., Hatzistavrou T, Crack P., Busciglio J, Kola I, Hertzog P. 2003. Overexpression of the chromosome 21 transcription factor Ets2 induces neuronal apoptosis. Neurobiol. Dis. 14:349–356.

Wray G a. 2007. The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8:206–16.

Wu C, Rolfe P, Gifford D, Fink G. 2010. Control of Transcription by Cell Size. PLoS Biol. 8:e1000523.

Yoshida K, Toki T, Okuno Y, Kanezaki R, Shiraishi Y, Sato-Otsubo A, Sanada M, Park M, Terui K, Suzuki H, et al. 2013. The landscape of somatic mutations in Down syndrome-related myeloid disorders. Nat. Genet. 45:1293–9.

Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang C-Z, Wala J, Mermel CH, et al. 2013. Pan-cancer patterns of somatic copy number alteration. Nat. Genet. 45:1134–1140.

Zeyl C, Vanderford T, Carter M. 2003. An evolutionary advantage of haploidy in large yeast populations. Science 299:555–558.

Zhang H, Zeidler AFB, Song W, Puccia CM, Malc E, Greenwell PW, Mieczkowski PA, Petes TD, Argueso JL. 2013. Gene copy-number variation in haploid and diploid strains of the yeast Saccharomyces cerevisiae. Genetics 193:785–801.

Zhang Y, Liu T, Meyer C a, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9:R137.

Zhu YO, Sherlock G, Petrov DA. 2016. Whole Genome Analysis of 132 Clinical Saccharomyces cerevisiae Strains Reveals Extensive Ploidy Variation. G3 Genes, Genomes, Genet. 6:2421–2434.

8 ADDITIONAL METHODS

8.1 Haploid, diploid, and tetraploid ancestral strain construction

All S. cerevisiae strains used in this study were isogenic to PY3295 (BY4741, S288c *MATa his3Δ leu2Δ met15Δ ura3Δ*) and are listed in (Selmecki et al. 2015). The strategy used to generate the isogenic ploidy series is illustrated in Figure 8-1 and genotypes of key intermediates are indicated. The *CFP* and *YFP* ancestors were derived from the haploid strain PY5997 (*matΔ::pSTE5-ura3::TRP1, ste4Δ::Hygro^R, trp1::Nat^R*). Isogenic strains with either the *CFP* or *YFP* cassettes at the *TRP1* locus (chromosome IV) were generated (PY5998 and PY5999) as follows: the *pGAL1–CFP–tADH–SpHIS5 or pGAL1–YFP–tADH–SpHIS5* cassette was PCR amplified from plasmid pB2694 (*CFP*) or pB2697 (*YFP*), respectively, with primers delTRPGFP5' (5'-TATTGAGCACGTGAGTATACGTGATTAAG CACACAAAGGCAGCTTGGAGTGCAGG TCGACGGATCCCCGGG-3') and delTRPGFP3' ('9- GAACGTGCACTGAGTAGTATG TTGCAGTCTTTTGGAAATACGAGTCGAATTCGAGCTCGTTTAAAC-3') and transformed into PY5997 at the *TRP1* locus. The haploid ancestor strains expressing *CFP* (PY5998) or *YFP* (PY5999) were confirmed by PCR and fluorescence microscopy. The haploid ancestors were modified to become mating competent by transformation with plasmids PB2647 (*LEU2-STE4*) and PB2649 (*URA3-STE4-MATa*). Diploid zygotes were selected on -Ura -Leu plates, and then colony purified on YPD plates to allow plasmid loss. Diploid chromosome content was confirmed by flow cytometry and aCGH, and strains PY6006, PY6008, PY6014, and PY6022 were selected. The diploid ancestors were made mating competent by transformation with plasmids PB2647 and PB2649. Tetraploid zygotes were pulled onto YPD plates using a micromanipulator, and

after 2 days growth at 30 uC the ploidy of each zygote was determined by flow cytometry and aCGH.

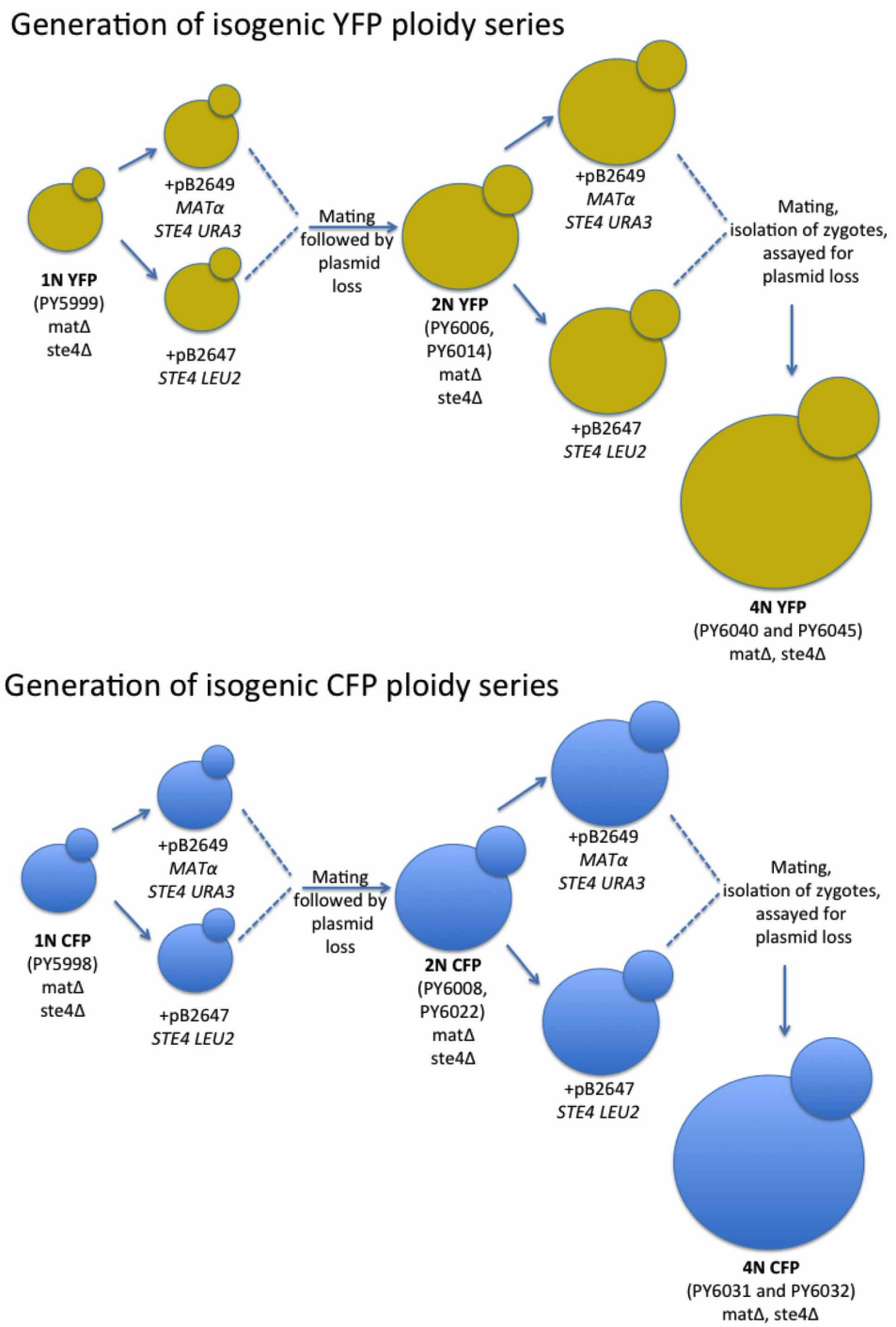


Figure 8-1 Schematic representation of the construction of isogenic haploid, diploid, and tetraploid strains used in this study. Relevant strain numbers are indicated for the *CFP*- and *YFP*-containing ancestors.

8.2 Plasmid Construction

To construct plasmids for the inducible expression of either *CFP* or *YFP*, the galactose-inducible *GAL1* promoter was sub-cloned into the *YFP* plasmid PB1500 and the *CFP* plasmid PB2452. These plasmids were derived from the green fluorescent protein (GFP) tagging plasmid generated in (Longtine et al. 1998). Both plasmids contained the ADH gene terminator (*tADH*) after the *YFP* or *CFP* gene and the sequence of the *SpHIS5* gene of *Schizosaccharomyes pombe* as a selectable marker. Plasmids PB1500 and PB2452 were digested with BamHI and PacI to introduce the *pGAL* promoter, 461 base pairs (bp) upstream of the start codon of *GAL1* (Mumberg et al. 1995), which was amplified using the primers *pGAL1* BamHI 5' (5'-ACGGATCCCCGGGTTGAAGTACGGATTAGAAGCCGCCGAG-3') and *pGAL1* PacI 3' (5'-CGTTAATTAATATAGTTTTTTCTCCTTGACGTTAAAG-3'). Site-directed mutagenesis (Quick Change Mutagenesis Kit, Stratagene) was used to introduce an ATG translation start codon to the YFP and CFP genes (using the GAPATGpFA6 primer, 5'-CAA TCAATCAATCAATCATCACATAAATTAATTAAATGAGTAAAGGAGAAG AACTTTTCACTGGAGTTGTC-3'). The resulting plasmids PB2694 and PB2697 contained the cassette *pGAL1–CFP–tADH–SpHIS5* and *pGAL1–YFP–tADH– SpHIS5*, respectively.

PB2314 was used to delete the MAT locus as previously described (Storchová et al. 2006). PB1308 was used to perform a URA3 to TRP1 marker swap, as previously described (Cross 1997). PB1640 (hphMX4, (Goldstein and McCusker 1999)) was used for PCR-mediated deletion of *STE4*. PB2647 (*STE4– LEU2*) was used to restore mating

competency and was constructed by amplifying the *STE4* gene with primers *STE4* P BamHI 5' (5'-CCGGATTCTTGTAGCCCTG TTAGGTTTACC-3') and *STE4* T BamHI 3' (5'-CCGGATTCCAATACATAAG GACGAGCCAGTG-3'), and cloning it into pRS315. PB2649 (*STE4 URA3 CEN MATa*) was also used to restore mating competency, and was constructed by sub cloning the *STE4* fragment from PB2647 (digested with SmaI and NotI) into PB2577 (*MATa URA3 CEN*, digested with SmaI and NotI).

8.3 Experimental evolution study

The experimental evolution study was performed by Anna Selmecki. All Saccharomyces cerevisiae strains used in the experimental evolution study were in the S288c background. The isogenic ploidy series was generated in a *matΔ ste4Δ* background to eliminate mating and meiosis during the course of the experiment. Either a *pGAL–CFP* or a *pGAL–YFP* construct was integrated at the *TRP1* locus near the chromosome IV centromere in a haploid strain (PY5998 and PY5999, respectively). These haploid strains were used to generate isogenic diploids, from which isogenic tetraploids were then derived. This procedure ensured that all copies of chromosome IV had the capacity to express the inducible fluorescent marker even if the strains became aneuploid. Mating-competent haploids were generated from the *matΔ ste4Δ* ancestor, PY5998, by transformation with either plasmid PB2647 (*CEN– LEU2–STE4*) or PB2648 (*CEN–URA3–STE4–Matα*). Zygotes from mating-competent haploids were isolated by micromanipulation to obtain diploid *CFP* ancestors (PY6008 and PY6022). Similarly, zygotes from mating-competent diploids were isolated by micro-manipulation to obtain tetraploid *CFP* ancestors (PY6031 and PY6032). The same mating scheme was

performed for the *YFP* lineage starting with PY5999 to generate diploid *YFP* (PY6006 and PY6014) and tetraploid *YFP* (PY6040 and PY6045) ancestors.

The ancestor strains were grown to saturation from the -80°C stock, in Synthetic Complete + 2% glucose. The cell density of each ancestor was determined using a haemocytometer and an automated cell counter (Vi-Cell-XR from Beckman Coulter). An equal number of YFP and CFP cells of the same ploidy were diluted into fresh Synthetic Complete + 2% raffinose medium, and combined into a single tube for an initial concentration of $1x10^5$ cells per millilitre. The 50:50 YFP:CFP culture was distributed equally into the wells of a 96-deep-well plate (1 ml per well, U-bottom block plate from Qiagen). Seven or eight wells were not inoculated, to detect cross-well contamination during the experiment. The plates were covered with 'breathe-EASIER' tape (Electron Microscopy Science) and incubated at 30°C on a 96-well plate shaker (Union Scientific). Two plates of haploid and three plates of diploid and tetraploid cells were analyzed, representing 173 parallel haploid evolutions, 264 parallel diploid evolutions, and 265 parallel tetraploid evolutions.

At 24 h intervals, the cells were resuspended (by pipetting) and diluted into fresh Synthetic Complete + 2% raffinose medium. The dilution factor was determined for each ploidy type based on the initial strain fitness to maintain an equivalent population size, as reported previously (Hegreness 2006). The number of cells transferred each day was calculated by counting the number of cells in ten replicate wells of each ploidy before and after dilution with an automated cell counter (Vi-Cell-XR from Beckman Coulter), and averaged across 3 consecutive days. The dilution factor for the haploid, diploid, and

tetraploid experiments was 1/100, 1/50, and 1/33, respectively. This corresponds to 6.64, 5.64, and 5.04 generations per day (Hegreness 2006).

The number of *CFP* and *YFP* cells in each population was measured at the same time each day. First, expression of the fluorescent proteins was induced by transferring 10 ml of the overnight culture into 200 ml Synthetic Complete + 2% galactose medium for 4 h at 30°C. The number of *CFP*- or *YFP*-expressing cells was determined using a BD LSRII flow cytometer high-throughput plate reader (10,000 cells were analyzed from each well). Pacific Blue and FITC filters were used to detect *CFP* and *YFP*, respectively. All experiments were passaged for 250 generations, but daily acquisition of *CFP:YFP* ratios was not always continued to the 250th generation.

To ensure that the flow cytometer measurement and the galactose induction of *CFP* and *YFP* was an accurate reflection of the size of these populations, the ratio of *CFP:YFP* cells was determined by both flow cytometry and microscopy, and the ratio was determined both before and after galactose induction. To do this, we combined overnight cultures of the 1N, 2N, and 4N ancestor *CFP* and *YFP* strains at three different ratios (nine populations in total) and analyzed the ratios in two ways. First, for an aliquot of the mixture, we induced the expression of the fluorescent proteins with 2% galactose for 4 h and analyzed 10,000 cells using flow cytometry. In parallel, we also added 2% galactose for 4 h and then counted about 300 cells by fluorescence microscopy. Finally, to ensure that the induction with 2% galactose did not alter the *CFP:YFP* ratio, a portion of the population was used to determine the number of *CFP* and *YFP* cells in the population before adding galactose to the medium. To do this, cells from each population were struck for single colonies on YPD plates for 2 days. Ninety-

six colonies were chosen randomly from each plate and added to a single well of a 96-well plate containing Synthetic Complete + 2% galactose. The fluorescence of each colony was determined by flow cytometry, and the percentage *YFP* of the initial population was determined. There was a strong correlation between the percentage *YFP*-expressing cells obtained from all three measurements: including the flow cytometer and fluorescence microscopy (Pearson correlation coefficient = 0.979), and both before and after galactose induction (Pearson correlation coefficient = 0.985).

Finally, frozen stocks of the evolution experiments were made at 3- to 4-day intervals throughout the experiment. At the end of each experiment, single colony clones were isolated and used for competitive fitness assays, flow cytometry analysis of ploidy, and preparation of DNA for aCGH and whole genome sequencing.

9 MUTATIONS IDENTIFIED IN THE EVOLVED CLONES BY WGS

Evolved Clone	Position		Allele Change	ORF	Gene Name	Type	Amino Acid	Sequencing Method
1N_101	chrIV	113135	G->T	YDL194W	SNF3	NonSYN	G519C	HiSeq2500
1N_101	chrXVI	793611	G->A	YPR130C		NonSYN	T56M	HiSeq2500
1N_102	chrIV	1014640	GT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_103	chrIV	590687	G->C	YDR072C	IPT1	NonSYN	C219W	HiSeq2500
1N_104	chrIV	1014640	GT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_104	chrXIII	235275	G->T	YML018C		SYN	N/A	HiSeq2500
1N_104	chrXVI	421710	C->T	YPL070W	MUK1	NonSYN	Q256Stop	HiSeq2500
1N_105	chrIV	1015664	G->GT	YDR277C	MTH1	FrameShift		HiSeq2500
1N_105	chrIV	1015670	C->G	YDR277C	MTH1	SYN	N/A	HiSeq2500
1N_105	chrXV	133333	T->A	YOL098C		NonSYN	N836Y	HiSeq2500
1N_105	chrXV	302669	C->T	YOL013C	HRD1	NonSYN	W123Stop	HiSeq2500
1N_105	chrXV	509038	T->C	YOR098C	NUP1	SYN	N/A	HiSeq2500
1N_105	chrXV	620188	G->C	YOR153W	PDR5	SYN	N/A	HiSeq2500
1N_106	chrIV	1014640	GT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_106	chrVIII	366177	C->T	YHR131C		SYN	N/A	HiSeq2500
1N_108	chrIV	1014640	GT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_108	chrXIV	267819	G->T	YNL197C	WHI3	SYN	N/A	HiSeq2500
1N_109	chrIV	1014703	T->A	YDR277C	MTH1	NonSYN	K333Stop	HiSeq2500
1N_110	chrIV	1015493	GTT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_111	chrIV	1014599	CT->C	YDR277C	MTH1	FrameShift		HiSeq2500
1N_112	chrIV	1015302	GA->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_112	chrXI	87233	C->G	YKL188C	PXA2	NonSYN	G520A	HiSeq2500
1N_112	chrXII	1051603	T->C	YLR454W	FMP27	SYN	N/A	HiSeq2500
1N_113	chrXIII	410221	A->T	YMR070W	MOT3	NonSYN	R357Stop	HiSeq2500
1N_114	chrXIII	409953	TC->T	YMR070W	MOT3	FrameShift		HiSeq2500
1N_116	chrIV	1014640	GT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_117	chrIV	590649	A->T	YDR072C	IPT1	NonSYN	V232D	HiSeq2500
1N_118	chrIV	113043	C->A	YDL194W	SNF3	NonSYN	A488D	HiSeq2500
1N_118	chrIX	308203	C->T	YIL025C		NonSYN	R124Q	HiSeq2500
1N_120	chrX	375167	G->T	YJL038C	LOH1	NonSYN	T201K	HiSeq2500
1N_120	chrXIII	410332	A->T	YMR070W	MOT3	NonSYN	K394Stop	HiSeq2500
1N_120	chrXVI	287670	C->G	YPL140C	MKK2	NonSYN	R455T	HiSeq2500
1N_121	chrIV	112989	T->A	YDL194W	SNF3	NonSYN	V470D	HiSeq2500
1N_121	chrXI	100694	A->C	YKL182W	FAS1	SYN	N/A	HiSeq2500
1N_122	chrIV	1014640	GT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_124	chrIV	1014640	GT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_124	chrV	241395	A->T	YER045C	ACA1	NonSYN	Y36N	HiSeq2500
1N_124	chrVII	92569	A->G	YGL211W	NCS6	SYN	N/A	HiSeq2500
1N_124	chrIX	173470	G->C	YIL102C-A		NonSYN	S40R	HiSeq2500
1N_127	chrII	528787	T->C	YBR142W	MAK5	SYN	N/A	HiSeq2500
1N_127	chrIV	1014876	G->T	YDR277C	MTH1	NonSYN	A275D	HiSeq2500
1N_127	chrX	29781	G->A	Intergenic	N/A	N/A	N/A	HiSeq2500
1N_127	chrXIV	361234	A->G	YNL139C	THO2	NonSYN	S1496P	HiSeq2500
1N_128	chrIV	1014640	GT->G	YDR277C	MTH1	FrameShift		HiSeq2500
1N_128	chrXIII	400289	T->A	YMR065W	KAR5	NonSYN	F197I	HiSeq2500
1N_131	chrIV	272666	G->A	YDL105W	NSE4	NonSYN	G93D	MiSeq
1N_131	chrIV	1014641	GT->G	YDR277C	MTH1	Frameshirt	I353fs	MiSeq

Evolved Clone	Position		Allele Change	ORF	Gene Name	Type	Amino Acid	Sequencing Method
1N_131	chrXII	558097	T->G	YLR207W	HRD3	Nonsense	T436Stop	MiSeq
1N_131	chrXIII	247081	G->T	YML011C	RAD33	SYN	N/A	MiSeq
1N_131	chrXIII	787006	G->A	YMR259C	TRM732	NonSYN	S626N	MiSeq
1N_132	chrII	760672	T->C	YBR278W	DPB3	NonSYN	D128G	MiSeq
1N_132	chrIV	591299	T->A	YDR072C	IPT1	Nonsense	T15Stop	MiSeq
1N_132	chrXVI	788744	G->T	YPR125W	YLH47	NonSYN	P263Q	MiSeq
2N_202	chrIV	112895	G->A	YDL194W	SNF3	NonSYN	G439R	HiSeq2500
2N_202	chrIV	1019186	T->C	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_202	chrVII	424078	T->C	YGL039W		NonSYN	S38P	HiSeq2500
2N_203	chrIV	111920	G->T	YDL194W	SNF3	NonSYN	D114Y	HiSeq2500
2N_204	chrII	213717	G->C	YBL007C	SLA1	NonSYN	Q885E	HiSeq2500
2N_204	chrII	523068	C->A	YBR140C	IRA1	SYN	N/A	HiSeq2500
2N_204	chrIV	213995	C->T	YDL138W	RGT2	NonSYN	S215F	HiSeq2500
2N_204	chrX	379508	C->A	YJL036W	SNX4	SYN	N/A	HiSeq2500
2N_204	chrXV	325765	G->A	YOL001W	PHO80	SYN	N/A	HiSeq2500
2N_205	chrXVI	192641	A->G	YPL188W	POS5	NonSYN	N413D	HiSeq2500
2N_206	chrIV	111920	G->T	YDL194W	SNF3	NonSYN	D114Y	HiSeq2500
2N_206	chrV	89212	A->T	YEL032W	MCM3	NonSYN	K759M	HiSeq2500
2N_206	chrXV	658941	C->A	YOR174W	MED4	NonSYN	P65Q	HiSeq2500
2N_207	chrIV	112952	T->C	YDL194W	SNF3	NonSYN	F458L	HiSeq2500
2N_207	chrIV	407030	C->G	YDL025C	RTK1	NonSYN	D59H	HiSeq2500
2N_208	chrIV	112238	C->G	YDL194W	SNF3	NonSYN	Q220E	HiSeq2500
2N_208	chrVII	378678	A->G	YGL066W	SGF73	NonSYN	E356G	HiSeq2500
2N_208	chrXV	544522	G->T	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_209	chrIV	111920	G->T	YDL194W	SNF3	NonSYN	D114Y	HiSeq2500
2N_209	chrXV	536313	T->A	YOR113W	AZF1	SYN	N/A	HiSeq2500
2N_210	chrIV	111882	G->C	YDL194W	SNF3	NonSYN	G101A	HiSeq2500
2N_210	chrV	452165	C->T	YER140W	EMP65	SYN	N/A	HiSeq2500
2N_211	chrII	94117	C->A	YBL067C	UBP13	NonSYN	V590F	HiSeq2500
2N_211	chrIV	111920	G->T	YDL194W	SNF3	NonSYN	D114Y	HiSeq2500
2N_211	chrVII	200627	A->T	YGL160W	AIM14	NonSYN	I21F	HiSeq2500
2N_212	NONE Found	NONE Found	NONE Found	NONE Found	N/A	NONE Found	NONE Found	HiSeq2500
2N_213	chrIV	111882	G->C	YDL194W	SNF3	NonSYN	G101A	HiSeq2500
2N_213	chrXIII	499722	C->A	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_214	chrIV	112221	C->A	YDL194W	SNF3	NonSYN	A214E	HiSeq2500
2N_214	chrXII	950618	G->T	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_214	chrXIV	637715	C->A	YNR006W	VPS27	NonSYN	H243N	HiSeq2500
2N_218	chrXI	192209	A->T	YKL133C		NonSYN	Y287Stop	HiSeq2500
2N_218	chrXIII	406673	G->T	YMR068W	AVO2	NonSYN	C124F	HiSeq2500
2N_219	chrIV	112954	C->A	YDL194W	SNF3	NonSYN	F458L	HiSeq2500
2N_219	chrIV	771695	T->A	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_221	chrIV	111920	G->T	YDL194W	SNF3	NonSYN	D114Y	HiSeq2500
2N_221	chrVIII	48667	G->C	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_221	chrVIII	489018	T->G	YHR194W	MDM31	NonSYN	F121C	HiSeq2500
2N_221	chrXIV	258430	C->T	YNL204C	SPS18	SYN	N/A	HiSeq2500
2N_222	chrIV	112830	G->C	YDL194W	SNF3	NonSYN	R417T	HiSeq2500
2N_222	chrIV	1226409	G->C	YDR375C	BCS1	NonSYN	L41V	HiSeq2500
2N_222	chrXIII	237806	C->A	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_222	chrXVI	448343	C->G	YPL058C	PDR12	NonSYN	G678R	HiSeq2500
2N_223	chrII	42494	C->T	YBL097W	BRN1	NonSYN	A556V	HiSeq2500

Evolved Clone	Position		Allele Change	ORF	Gene Name	Type	Amino Acid	Sequencing Method
2N_223	chrIV	111913	T->G	YDL194W	SNF3	NonSYN	F111L	HiSeq2500
2N_223	chrIV	800105	G->C	YDR170C	SEC7	NonSYN	P706A	HiSeq2500
2N_223	chrX	153262	A->G	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_223	chrXI	161378	C->T	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_223	chrXII	41954	G->A	YLL048C	YBT1	SYN	N/A	HiSeq2500
2N_223	chrXV	897608	G->T	YOR310C	NOP58	SYN	N/A	HiSeq2500
2N_224	chrIV	111920	G->T	YDL194W	SNF3	NonSYN	D114Y	HiSeq2500
2N_224	chrVI	120303	C->T	YFL008W	SMC1	NonSYN	Q294Stop	HiSeq2500
2N_224	chrXII	99093	C->A	YLL022C	HIF1	NonSYN	D370Y	HiSeq2500
2N_224	chrXIV	130327	G->A	YNL271C	BNI1	SYN	N/A	HiSeq2500
2N_224	chrXV	1074817	G->T	YOR389W		NonSYN	E203Stop	HiSeq2500
2N_225	chrIV	111920	G->T	YDL194W	SNF3	NonSYN	D114Y	HiSeq2500
2N_225	chrIV	1107015	G->A	YDR320C	SWA2	NonSYN	S362F	HiSeq2500
2N_226	chrIV	111920	G->T	YDL194W	SNF3	NonSYN	D114Y	HiSeq2500
2N_226	chrV	88495	T->G	YEL032W	MCM3	NonSYN	F520C	HiSeq2500
2N_226	chrX	589861	A->C	YJR089W	BIR1	NonSYN	N718H	HiSeq2500
2N_227	chrIV	112954	C->A	YDL194W	SNF3	NonSYN	F458L	HiSeq2500
2N_227	chrIV	244559	T->C	YDL122W	UBP1	NonSYN	S670P	HiSeq2500
2N_227	chrVII	69728	T->G	YGL227W	VID30	NonSYN	Y20D	HiSeq2500
2N_227	chrX	135085	G->A	YJL153C	INO1	NonSYN	A282V	HiSeq2500
2N_227	chrXV	326628	AT->A	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_227	chrXV	581532	T->A	Intergenic	N/A	N/A	N/A	HiSeq2500
2N_227	chrXVI	221645	A->G	YPL174C	NIP100	SYN	N/A	HiSeq2500
2N_232	chrXIII	746160	T->A	Intergenic	N/A	N/A	N/A	SOLiD
2N_233	chrI	103998	T->A	YAL024C	LTE1	NonSYN	S626C	SOLiD
2N_233	chrIV	112896	G->A	YDL194W	SNF3	NonSYN	G439E	SOLiD
4N_304	chrIV	112896	G->T	YDL194W	SNF3	NonSYN	G439V	HiSeq2500
4N_304	chrV	79234	A->T	YEL040W	UTR2	SYN	N/A	HiSeq2500
4N_304	chrXI	3532	C->G	YKL222C		NonSYN	S697T	HiSeq2500
4N_304	chrXIII	23797	G->C	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_304	chrXIII	653210	T->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_304	chrXV	106555	G->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_305	chrIV	57594	T->C	YDL223C	HBT1	NonSYN	E938G	HiSeq2500
4N_305	chrIV	112817	G->A	YDL194W	SNF3	NonSYN	E413K	HiSeq2500
4N_305	chrIV	415931	G->A	YDL020C	RPN4	NonSYN	T259M	HiSeq2500
4N_305	chrVI	34264	C->G	YFL050C	ALR2	NonSYN	G529R	HiSeq2500
4N_305	chrXIII	334443	G->C	YMR031C	EIS1	SYN	N/A	HiSeq2500
4N_305	chrXIV	589741	C->G	YNL023C	FAP1	NonSYN	Q474H	HiSeq2500
4N_305	chrXVI	122596	C->A	YPL226W	NEW1	NonSYN	S277Y	HiSeq2500
4N_305	chrXVI	215384	T->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_306	chrX	712087	A->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_306	chrXI	551081	T->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_306	chrXIII	504322	G->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_307	chrIV	112266	G->A	YDL194W	SNF3	NonSYN	R229K	HiSeq2500
4N_307	chrXVI	110144	G->T	YPL231W	FAS2	NonSYN	G498V	HiSeq2500
4N_307	chrXVI	341104	G->A	YPL110C	GDE1	NonSYN	A1212V	HiSeq2500
4N_309	chrIII	152835	A->T	YCR019W	MAK32	NonSYN	M1L	HiSeq2500
4N_309	chrIV	112734	C->G	YDL194W	SNF3	NonSYN	T385R	HiSeq2500
4N_309	chrIV	779067	G->A	YDR161W		SYN	N/A	HiSeq2500
4N_309	chrVIII	509672	T->C	YHR205W	SCH9	NonSYN	S104P	HiSeq2500
4N_309	chrVIII	554783	C->A	YHR216W	IMD2	NonSYN	P130T	HiSeq2500

Evolved Clone	Position		Allele Change	ORF	Gene Name	Type	Amino Acid	Sequencing Method
4N_310	chrIV	112268	G->A	YDL194W	SNF3	NonSYN	G230S	HiSeq2500
4N_310	chrVII	103390	G->A	YGL206C	CHC1	SYN	N/A	HiSeq2500
4N_310	chrVIII	30115	C->T	YHL035C	VMR1	SYN	N/A	HiSeq2500
4N_310	chrXIII	315766	C->A	YMR020W	FMS1	NonSYN	H131N	HiSeq2500
4N_310	chrXIII	463918	A->AT	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_310	chrXIII	756312	C->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_310	chrXIV	380225	A->C	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_311	chrIV	112251	C->G	YDL194W	SNF3	NonSYN	T224R	HiSeq2500
4N_311	chrVI	15959	C->T	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_311	chrVII	617090	T->G	YGR062C	COX18	NonSYN	M65L	HiSeq2500
4N_311	chrIX	71149	C->T	YIL147C	SLN1	NonSYN	G769R	HiSeq2500
4N_311	chrXIV	137168	C->T	YNL270C	ALP1	SYN	N/A	HiSeq2500
4N_312	chrVII	84042	C->A	YGL218W		NonSYN	P132T	HiSeq2500
4N_313	chrIV	112896	G->T	YDL194W	SNF3	NonSYN	G439V	HiSeq2500
4N_313	chrIV	488849	AACCACTGG->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_313	chrV	79234	A->T	YEL040W	UTR2	SYN	N/A	HiSeq2500
4N_313	chrV	304535	A->T	YER073W	ALD5	NonSYN	N170I	HiSeq2500
4N_313	chrXI	296675	G->T	YKL073W	LHS1	NonSYN	G201V	HiSeq2500
4N_313	chrXIII	23797	G->C	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_313	chrXIII	653210	T->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_313	chrXV	811463	C->A	YOR257W	CDC31	NonSYN	N152K	HiSeq2500
4N_314	chrIII	84937	G->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_314	chrIV	229797	C->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_314	chrVI	256200	G->C	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_314	chrVIII	186070	C->A	YHR039C	MSC7	NonSYN	V245F	HiSeq2500
4N_314	chrIX	277539	C->CTA	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_314	chrXII	330455	C->T	YLR095C	IOC2	NonSYN	M554I	HiSeq2500
4N_314	chrXII	820836	G->C	YLR345W		NonSYN	W109S	HiSeq2500
4N_314	chrXV	881004	C->G	YOR301W	RAX1	NonSYN	F14L	HiSeq2500
4N_315	chrIV	214604	G->A	YDL138W	RGT2	NonSYN	G418D	HiSeq2500
4N_315	chrIV	1389304	G->A	YDR464W	SPP41	NonSYN	R147Q	HiSeq2500
4N_315	chrVI	33920	C->T	YFL050C	ALR2	SYN	N/A	HiSeq2500
4N_315	chrXII	667355	G->A	YLR260W	LCB5	NonSYN	E504K	HiSeq2500
4N_316	chrIV	214594	G->A	YDL138W	RGT2	NonSYN	D415N	HiSeq2500
4N_319	chrVIII	488371	C->CA	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_319	chrXII	322378	G->T	YLR091W	GEP5	SYN	N/A	HiSeq2500
4N_319	chrXIII	211720	G->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_320	chrIV	1154143	T->C	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_320	chrIV	1154155	A->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_320	chrIV	1154200	G->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_320	chrVII	176888	G->T	YGL173C	XRN1	NonSYN	L1078I	HiSeq2500
4N_321	chrII	667548	G->A	YBR222C	PCS60	NonSYN	P267S	HiSeq2500
4N_321	chrVIII	72849	C->A	YHL016C	DUR3	NonSYN	G465C	HiSeq2500
4N_321	chrIX	277539	C->CTA	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_322	chrIV	112982	G->C	YDL194W	SNF3	NonSYN	G468R	HiSeq2500
4N_322	chrVII	328015	A->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_322	chrXV	1013114	G->GA	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_322	chrXVI	238978	A->T	YPL165C	SET6	SYN	N/A	HiSeq2500
4N_322	chrXVI	868272	G->A	YPR162C	ORC4	NonSYN	P10L	HiSeq2500
4N_323	chrII	92822	C->G	YBL068W	PRS4	NonSYN	I137M	HiSeq2500

Evolved Clone	Position		Allele Change	ORF	Gene Name	Type	Amino Acid	Sequencing Method
4N_323	chrIV	112251	C->G	YDL194W	SNF3	NonSYN	T224R	HiSeq2500
4N_323	chrIV	617159	C->T	YDR086C	SSS1	NonSYN	A4T	HiSeq2500
4N_323	chrXII	856265	T->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_324	chrX	338725	G->C	YJL052W	TDH1	NonSYN	C154S	HiSeq2500
4N_324	chrXI	366773	C->G	YKL038W	RGT1	NonSYN	S509Stop	HiSeq2500
4N_324	chrXII	127882	C->G	YLL011W	SOF1	NonSYN	P121A	HiSeq2500
4N_327	chrIV	1014738	C->A	YDR277C	MTH1	NonSYN	C321F	HiSeq2500
4N_327	chrVI	92904	G->T	YFL023W	BUD27	NonSYN	A641S	HiSeq2500
4N_328	chrII	804908	C->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_328	chrIV	1154247	G->A	YDR342C	HXT7	SYN	N/A	HiSeq2500
4N_328	chrV	465825	A->C	YER148W	SPT15	SYN	N/A	HiSeq2500
4N_328	chrX	384974	G->C	YJL033W	HCA4	NonSYN	G382R	HiSeq2500
4N_328	chrXII	304261	T->A	YLR086W	SMC4	NonSYN	L673Stop	HiSeq2500
4N_328	chrXVI	44175	A->C	YPL264C		NonSYN	V57G	HiSeq2500
4N_329	chrII	278810	GT->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_329	chrV	465825	A->C	YER148W	SPT15	SYN	N/A	HiSeq2500
4N_329	chrVII	833536	G->A	YGR168C		NonSYN	R318W	HiSeq2500
4N_329	chrXI	61897	C->A	YKL203C	TOR2	NonSYN	M488I	HiSeq2500
4N_329	chrXV	961779	G->A	YOR341W	RPA190	NonSYN	V266I	HiSeq2500
4N_331	chrI	113939	C->T	YAL020C	ATS1	SYN	N/A	HiSeq2500
4N_331	chrIV	112988	G->T	YDL194W	SNF3	NonSYN	V470F	HiSeq2500
4N_331	chrXVI	309088	G->A	YPL127C	HHO1	SYN	N/A	HiSeq2500
4N_332	chrIV	1154143	T->C	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_332	chrIV	1154155	A->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_332	chrIV	1154200	G->A	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_332	chrV	505900	C->T	YER164W	CHD1	NonSYN	Q172Stop	HiSeq2500
4N_332	chrVII	80102	G->A	YGL223C	COG1	NonSYN	T88I	HiSeq2500
4N_332	chrIX	248488	C->G	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_332	chrXI	162068	A->C	YKL154W	SRP102	NonSYN	K155Q	HiSeq2500
4N_332	chrXVI	243118	G->T	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_332	chrXVI	341131	T->G	YPL110C	GDE1	NonSYN	D1203A	HiSeq2500
4N_333	chrIII	50511	C->T	YCL041C		SYN	N/A	HiSeq2500
4N_333	chrVIII	234713	A->C	Intergenic	N/A	N/A	N/A	HiSeq2500
4N_333	chrXVI	122727	T->G	YPL226W	NEW1	NonSYN	F321V	HiSeq2500
4N_334	chrV	465825	A->C	YER148W	SPT15	SYN	N/A	SOLiD
4N_334	chrVII	1029503	A->C	YGR270W	YTA7	NonSYN	M710L	SOLiD
4N_334	chrXI	61897	C->A	YKL203C	TOR2	NonSYN	M488I	SOLiD
4N_335	chrIV	609668	C->T	YDR081C	PDC2	NonSYN	V138I	SOLiD
4N_335	chrVI	133989	G->T	YFL004W	VTC2	NonSYN	A729S	SOLiD
4N_335	chrVII	1023892	G->A	YGR266W	N/A	NonSYN	A411T	SOLiD
4N_335	chrXV	332208	G->A	YOR003W	YSP3	SYN	N/A	SOLiD
4N_335	chrXVI	671171	C->A	YPR056W	TFB4	NonSYN	Q17K	SOLiD
4N_335	chrVIII	23427	A->C	YHL039W	EFM1	NonSYN	I550L	SOLiD
4N_336	chrVII	852082	T->A	YGR178C	PBP1	NonSYN	F380Y	SOLiD
4N_336	chrXI	103753	G->T	YKL182W	FAS1	NonSYN	E1026D	SOLiD
4N_336	chrVIII	375275	C->A	Intergenic	N/A	N/A	N/A	SOLiD
4N_337	chrX	722854	T->A	YJR153W	PGU1	NonSYN	I17N	SOLiD
4N_337	chrII	289179	G->T	Intergenic	N/A	N/A	N/A	SOLiD
4N_337	chrIV	995263	G->C	YDR263C	DIN7	NonSYN	T90S	SOLiD
4N_337	chrVII	892626	G->A	YGR197C	SNG1	NonSYN	A507V	SOLiD
4N_337	chrXIII	646177	C->T	YMR191W	SPG5	Nonsense	Q175Stop	SOLiD

Table 9-1 Mutations identified in the evolved clones

10 CHROMOSOME COPY NUMBER

Evolved Clone	Ploidy (FACS)	Chromosome																Method
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	
1N_101	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
1N_102	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_103	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_104	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	CGH & Illumina
1N_105	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_106	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_107	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	CGH
1N_108	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_109	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_110	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	CGH & Illumina
1N_111	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_112	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_113	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	CGH & Illumina
1N_114	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	CGH & Illumina
1N_116	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_117	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_118	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_120	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_121	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_122	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_123	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	CGH
1N_124	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_127	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_128	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Illumina
1N_131	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	CGH & Illumina
1N_132	1N	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	CGH & Illumina

Table 10-1 Chromosome copy number in evolved haploids

Evolved Clone	Ploidy (FACS)	Chromosome																Method
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	
2N_201	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	CGH
2N_202	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_203	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_204	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_205	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_206	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_207	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_208	2N	2	2	2	2	2.1	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_209	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_210	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_211	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_212	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	CGH & Illumina
2N_213	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_214	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_217	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	CGH
2N_218	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	CGH & Illumina
2N_219	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_221	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_222	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_223	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_224	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_225	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_226	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_227	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
2N_228	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	CGH
2N_230	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	CGH
2N_232	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	CGH & SOLiD
2N_233	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	CGH & SOLiD

Table 10-2 Chromosome copy number in the evolved diploids

Evolved Clone	Ploidy (FACS)	Chromosome																Method
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	
4N_301	4N-	4	4	4	3.5	4	4	4	4	4	4	4	4	4	4	4	4	CGH
4N_302	3N+/-	2	4	2	3	4	3	3	2	3	2	4	3	4	4	3	3	CGH
4N_304	3N+	3	3	3	3	3	3	3	4	3	3	3	4	4	3	3	3	CGH & Illumina
4N_305	3N+	3	3	4	3	3	3	3	3	4	3	3	3	4	4	3	3	CGH & Illumina
4N_306	3N+/-	3	2	2	2	3	2	3	3	4	3	2	3	3	3	2	2	Illumina
4N_309	4N-3N	3	4	4	3	3	3	4	4	4	3	3	4	4	4	3	3	Illumina
4N_310	3N+	3	3	4	3	3	3	3	3	4	3	3	3	4	4	3	3	CGH & Illumina
4N_313	3N+	4	3	3	3	3	3	3	4	3	3	3	4	4	3	3	3	CGH & Illumina
4N_314	3N-2N	3	2	2	2	3	2	3	3	2	3	2	3	3	3	2.5	2	Illumina
4N_315	3N+	3	3	4	3	3	3	3	3	4	3	3	3	4	4	3	3	CGH & Illumina
4N_316	3N+	3	3	4	3	3	3	3	3	4	3	3	3	4	4	3	3	CGH & Illumina
4N_317	4N+/-	3	4	3	3	4	4	4	4	3	4	4	4	5	4	3	3	CGH
4N_318	3N-	3	2	2	2	3	2	3	3	3	3	2	3	3	3	2	3	CGH
4N_319	3N+	3	4	3	3	3	3	4	4	3	3	3	4	4	4	3	3	CGH & Illumina
4N_320	3N+	3	4	4	3	3	3	3	3	3	3	4	3	4	3	3	3	CGH & Illumina
4N_321	2N+	2	2	2	2	3	2	3	3	2	2	2	3	3	3	2	2	CGH & Illumina
4N_322	3N+	3	3	4	3	3	3	3	3	4	3	3	3	4	4	3	3	Illumina
4N_323	4N-	3	4	4	4	4	4	3	4	4	4	4	4	4	4	3	3	Illumina
4N_327	3N+	3	3	3	3	3	3	3	3	4	3	3	4	4	4	3	3	Illumina
4N_328	4N+	4	4	4	4	4	4	4	4	4	4	4	4	5	4	4	4	Illumina
4N_329	3N+	3	3	4	3	3	3	3	3	4	3	3	3	4	4	3	3	CGH & Illumina
4N_330	3N+	3	3	3	3	3	3	3	4	3	3	3	4	4	3	3	3	CGH
4N_331	2N	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Illumina
4N_332	4N+	4	4	4	4	4	4	4	4	4	4	4	4	5	4	4	4	Illumina
4N_333	4N+	4	4	4	4	4	4	4	4	4	4	4	4	5	4	4	4	Illumina
4N_334	3N+	3	3	3	3	3	3	3	3	4	3	3	3	4	4	3	3	CGH & SOLiD
4N_335	4N+/-	4	4	4	3.5	4	4	4	4	4	4	4	4	5	3	4	4	CGH & SOLiD
4N_336	4N+	4	4	4	4	4	4	4	4	4	4	4	6	5	4	4	4	CGH & SOLiD
4N_337	4N	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	CGH & SOLiD
4N_338	3N+	3	3	4	3	3	3	3	3	3	3	3	3	4	3	3	3	CGH

Table 10-3 Chromosome copy number in evolved tetraploids

11 MUTATIONS BY TYPE

	Haploid genome content	SYN	Non-syn	intergenic	Stop	Frameshift	Hxt6/7	CNVs (+HXT6/7)	Frameshift/ Stop	total (+CNV)
1N_101	1	0	2	0	0	0	0	0	0	2
1N_102	1	0	0	0	0	1	0	0	1	1
1N_103	1	0	1	0	0	0	0	0	0	1
1N_104	1	1	0	0	1	1	0	0	2	3
1N_105	1	3	1	0	1	1	0	0	2	6
1N_106	1	1	0	0	0	1	0	0	1	2
1N_108	1	1	0	0	0	1	0	0	1	2
1N_109	1	0	0	0	1	0	0	0	1	1
1N_110	1	0	0	0	0	1	0	0	1	1
1N_111	1	0	0	0	0	1	0	0	1	1
1N_112	1	1	1	0	0	1	0	0	1	3
1N_113	1	0	0	0	1	0	0	0	1	1
1N_114	1	0	0	0	0	1	0	0	1	1
1N_116	1	0	0	0	0	1	0	0	1	1
1N_117	1	0	1	0	0	0	0	0	0	1
1N_118	1	0	2	0	0	0	0	0	0	2
1N_120	1	0	2	0	1	0	0	0	1	3
1N_121	1	1	1	0	0	0	0	0	0	2
1N_122	1	0	0	0	0	1	0	0	1	1
1N_124	1	1	2	0	0	1	0	0	1	4
1N_127	1	1	2	1	0	0	0	0	0	4
1N_128	1	0	1	0	0	1	0	0	1	2
1N_131	1	1	2	0	1	1	0	0	2	5
1N_132	1	0	2	0	1	0	1	1	1	4
Average	na	0.4583	0.8333	0.0417	0.2917	0.5833	0.0417	0.0417	0.8750	**2.2500**
SD	na	0.7059	0.8498	0.1998	0.4545	0.4930	0.1998	0.1998	0.5995	**1.4216**
SEM	na	0.1441	0.1735	0.0408	0.0928	0.1006	0.0408	0.0408	0.1224	**0.2902**

Table 11-1 Haploid mutations by type

	Haploid genome content	SYN	Non-syn	intergenic	Stop	Frameshift	Hxt6/7	CNV (+HXT6/7)	Frameshift/ Stop	total (+CNV)	adj. total (+CNV)
2N_202	2	0	1	1	0	0	0	0	0	2	1
2N_203	2	0	1	0	0	0	0	0	0	1	0.5
2N_204	2	3	2	0	0	0	0	0	0	5	2.5
2N_205	2	0	1	0	0	0	1	1	0	2	1
2N_206	2	0	3	0	0	0	0	0	0	3	1.5
2N_207	2	0	2	0	0	0	1	1	0	3	1.5
2N_208	2	0	2	1	0	0	0	1	0	4	2
2N_209	2	1	1	0	0	0	0	0	0	2	1
2N_210	2	1	1	0	0	0	0	0	0	2	1
2N_211	2	0	3	0	0	0	0	0	0	3	1.5
2N_212	2	0	0	0	0	0	1	1	0	1	0.5
2N_213	2	0	1	1	0	0	0	0	0	2	1
2N_214	2	0	1	2	0	0	0	0	0	3	1.5
2N_218	2	0	1	0	1	0	1	1	1	3	1.5
2N_219	2	0	1	1	0	0	0	0	0	2	1
2N_221	2	1	2	1	0	0	0	0	0	4	2
2N_222	2	0	3	1	0	0	0	0	0	4	2
2N_223	2	2	3	2	0	0	0	0	0	7	3.5
2N_224	2	1	2	0	2	0	0	0	2	5	2.5
2N_225	2	0	2	0	0	0	0	0	0	2	1
2N_226	2	0	3	0	0	0	0	0	0	3	1.5
2N_227	2	1	4	2	0	0	0	0	0	7	3.5
2N_232	2	0	0	1	0	0	1	1	0	2	1
2N_233	2	0	2	0	0	0	0	0	0	2	1
Average	na	0.417	1.750	0.542	0.125	0.000	0.208	0.250	0.125	**3.083**	**1.542**
SD	na	0.759	1.010	0.706	0.439	0.000	0.406	0.433	0.439	**1.579**	**0.789**
SEM	na	0.155	0.206	0.144	0.090	0.000	0.083	0.088	0.090	**0.322**	**0.161**

Table 11-2 Diploid mutations by type

	Haploid genome content	SYN	Non-syn	intergenic	Stop	Frameshift	Hxt6/7	CNV (+HXT6/7)	Frameshift Stop	total (+CNV)	Adj. total +CNV
4N_304	3.1	1	2	3	0	0	0	0	0	6	1.95
4N_305	3.1	1	6	1	0	0	0	0	0	8	2.62
4N_306	2.7	0	0	3	0	0	1	1	0	4	1.49
4N_307	3.6	0	3	0	0	0	0	0	0	3	0.82
4N_309	3.4	1	4	0	0	0	0	0	0	5	1.46
4N_310	3.0	2	2	3	0	0	0	0	0	7	2.30
4N_311	3.5	1	3	1	0	0	0	0	0	5	1.44
4N_312	3.8	0	1	0	0	0	1	1	0	2	0.52
4N_313	3.1	1	4	3	0	0	0	0	0	8	2.59
4N_314	2.4	0	4	4	0	0	0	0	0	8	3.27
4N_315	3.1	1	2	0	0	0	0	0	0	3	0.97
4N_316	3.0	0	1	0	0	0	0	0	0	1	0.33
4N_319	3.3	1	0	2	0	0	1	1	0	4	1.20
4N_320	3.1	0	1	3	0	0	1	1	0	5	1.60
4N_321	2.4	0	2	1	0	0	0	0	0	3	1.27
4N_322	3.0	1	2	2	0	0	0	0	0	5	1.64
4N_323	3.7	0	3	1	0	0	0	0	0	4	1.08
4N_324	2.4	0	2	0	1	0	0	0	1	3	1.24
4N_327	3.3	0	2	0	0	0	0	0	0	2	0.60
4N_328	3.8	2	2	1	1	0	1	1	1	7	1.84
4N_329	3.2	1	3	1	0	0	0	0	0	5	1.58
4N_331	1.9	2	1	0	0	0	0	0	0	3	1.58
4N_332	3.7	0	2	5	1	0	1	1	1	9	2.43
4N_333	3.7	1	1	1	0	0	1	1	0	4	1.08
4N_334	3.1	1	2	0	0	0	0	0	0	3	0.98
4N_335	3.8	1	5	0	0	0	1	2	0	8	2.09
4N_336	4.2	0	2	1	0	0	0	0	0	3	0.72
4N_337	3.7	0	3	1	1	0	1	1	1	6	1.61
Average	na	0.64	2.32	1.32	0.14	0.00	0.32	0.36	0.14	**4.79**	**1.51**
SD	na	0.67	1.36	1.39	0.35	0.00	0.47	0.55	0.35	**2.13**	**0.68**
SEM	na	0.13	0.26	0.26	0.07	0.00	0.09	0.10	0.07	**0.40**	**0.13**

Table 11-3 Tetraploid mutations by type

12 DIFFERENTIAL EXPRESSION ANALYSIS

Comparison	Up/Down	Number of Genes	Gene Name	Systematic Name	log2fold Change	Adjusted P-value
2Nv131	Up	10	YFR032C-B	YFR032C-B	1.6851	3.89E-02
			ARI1	YGL157W	1.8471	2.34E-02
			MIG2	YGL209W	3.5751	3.59E-09
			ZRT1	YGL255W	3.4939	2.16E-09
			HXT4	YHR092C	4.1865	9.32E-13
			YJL133C-A	YJL133C-A	2.2195	2.94E-03
			AHP1	YLR109W	1.6571	4.42E-02
			HXT2	YMR011W	2.0447	2.94E-03
			ZPS1	YOL154W	3.0811	8.38E-07
			YOR170W	YOR170W	2.3227	2.77E-02
	Down	40	ACH1	YBL015W	-1.9330	6.64E-03
			ECM13	YBL043W	-2.4885	1.02E-04
			PHO5	YBR093C	-3.7965	1.63E-10
			PHO89	YBR296C	-5.9317	1.16E-20
			ADY2	YCR010C	-3.0487	3.11E-05
			GIT1	YCR098C	-2.0539	1.68E-02
			YDR034W-B	YDR034W-B	-2.0895	3.99E-02
			PHM6	YDR281C	-2.2619	5.37E-03
			ARO10	YDR380W	-2.1540	2.85E-02
			GLC3	YEL011W	-2.3519	2.47E-04
			RGI1	YER067W	-2.9130	1.10E-06
			YGR067C	YGR067C	-1.8056	3.04E-02
			MGA1	YGR249W	-2.3143	9.49E-04
			SPL2	YHR136C	-4.3087	5.47E-09
			SIP4	YJL089W	-1.9186	4.36E-02
			INO1	YJL153C	-1.8284	1.48E-02
			SFC1	YJR095W	-2.3766	2.90E-03
			CWP1	YKL096W	-1.9839	5.06E-03
			KDX1	YKL161C	-1.8020	2.47E-02
			JEN1	YKL217W	-2.1123	2.05E-03
			SRL3	YKR091W	-2.5077	1.86E-04
			PCK1	YKR097W	-2.9014	2.56E-06
			IDP2	YLR174W	-1.9988	5.06E-03
			FBP1	YLR377C	-3.7139	4.63E-10
			PHO84	YML123C	-5.6281	2.53E-20
			YPK2	YMR104C	-1.9085	8.32E-03
			SPG4	YMR107W	-2.7510	5.22E-04

			SIP18	YMR175W	-1.9061	2.12E-02
			YMR206W	YMR206W	-1.7883	2.77E-02
			ADH2	YMR303C	-2.0736	2.84E-03
			MLS1	YNL117W	-1.9159	2.13E-02
			YNL194C	YNL194C	-3.6059	3.98E-07
			BAG7	YOR134W	-2.6079	1.23E-04
			GAC1	YOR178C	-2.0107	4.24E-03
			TPO4	YOR273C	-1.9461	6.24E-03
			MF(ALPHA)1	YPL187W	-2.6379	2.52E-03
			USV1	YPL230W	-1.7429	3.30E-02
			CSR2	YPR030W	-2.0829	2.58E-03
			OPT2	YPR194C	-2.9868	2.45E-05
			YPR195C	YPR195C	-2.6013	4.42E-02
2Nv132	Up	9	HXT6	YDR343C	2.5943	7.88E-05
			SIT1	YEL065W	2.0634	7.19E-03
			SSA4	YER103W	1.8348	4.32E-02
			BTN2	YGR142W	2.2912	1.40E-03
			RGI2	YIL057C	2.0971	7.19E-03
			YJL133C-A	YJL133C-A	2.0553	2.50E-02
			SFC1	YJR095W	2.1158	1.33E-02
			ZPS1	YOL154W	3.0519	3.19E-06
			FDH1	YOR388C	3.4858	1.67E-06
	Down	5	PHO5	YBR093C	-2.3542	8.73E-04
			PHO89	YBR296C	-2.5382	2.50E-04
			SPL2	YHR136C	-2.5414	4.38E-03
			CWP1	YKL096W	-3.2988	2.03E-07
			PHO84	YML123C	-3.1846	4.90E-07
2Nv232	Up	9	HXT6	YDR343C	3.5373	4.33E-09
			SIT1	YEL065W	2.2803	1.31E-03
			SSA4	YER103W	3.5539	4.33E-09
			BTN2	YGR142W	4.0458	3.29E-11
			RGI2	YIL057C	2.3128	1.31E-03
			SFC1	YJR095W	2.1988	8.42E-03
			FDH1	YOR388C	4.7959	9.28E-12
			MF(ALPHA)1	YPL187W	2.7006	3.25E-04
			HSP82	YPL240C	2.7678	1.50E-05
	Down	7	PHO5	YBR093C	-2.5720	1.28E-04
			PHO89	YBR296C	-3.0352	2.94E-06
			GIT1	YCR098C	-2.3120	1.16E-02
			GLC3	YEL011W	-2.0899	6.03E-03
			SPL2	YHR136C	-2.8097	1.29E-03

			CWP1	YKL096W	-2.8648	6.82E-06
			PHO84	YML123C	-3.5991	4.33E-09
2Nv233	Up	4	HXT3	YDR345C	2.8824	5.56E-05
			MIG2	YGL209W	3.0230	2.18E-06
			HXT4	YHR092C	4.6507	2.35E-15
			YJL133C-A	YJL133C-A	2.1115	1.82E-02
	Down	7	PHO5	YBR093C	-3.4889	6.77E-09
			PHO89	YBR296C	-5.6373	1.72E-19
			PHM6	YDR281C	-3.3272	4.18E-06
			ARO10	YDR380W	-3.0507	3.63E-04
			SPL2	YHR136C	-3.6795	1.04E-06
			PHO84	YML123C	-6.4446	1.23E-24
			GRE1	YPL223C	-2.3684	7.65E-03
2Nv334	Up	21	NHP6b	YBR089C-A	1.8604	1.70E-02
			YDR133C	YDR133C	2.2177	1.06E-03
			YDR433W	YDR433W	2.2991	6.22E-03
			SIT1	YEL065W	2.6101	4.23E-05
			SPI1	YER150W	1.9167	1.27E-02
			YFR032C-B	YFR032C-B	1.9495	8.90E-03
			ZRT1	YGL255W	2.3805	2.97E-04
			YGR025W	YGR025W	2.0749	5.31E-03
			YIR016W	YIR016W	1.6982	4.98E-02
			TIS11	YLR136C	2.0096	9.10E-03
			YET2	YMR040W	1.7612	4.98E-02
			ISF1	YMR081C	1.8331	1.85E-02
			HOR7	YMR251W-A	2.2956	5.91E-04
			COX7	YMR256C	1.7801	3.16E-02
			NCE103	YNL036W	2.1030	2.95E-03
			YNR014W	YNR014W	2.1770	4.49E-03
			HUB1	YNR032C-A	2.0818	6.09E-03
			ZPS1	YOL154W	1.9366	1.76E-02
			FIT2	YOR382W	2.6573	4.23E-05
			FIT3	YOR383C	2.7785	1.18E-05
			FRE5	YOR384W	2.6779	9.63E-05
	Down	11	PHO5	YBR093C	-4.1135	4.47E-12
			PHO89	YBR296C	-5.7606	4.40E-20
			GIT1	YCR098C	-1.9869	3.23E-02
			PHM6	YDR281C	-3.4585	2.14E-06
			COS12	YGL263W	-3.4451	8.16E-04
			ECL1	YGR146C	-2.8371	3.42E-05
			SPL2	YHR136C	-4.4747	3.00E-09

			YHR177W	YHR177W	-3.5369	5.91E-04
			PHO12	YHR215W	-2.9406	9.10E-03
			PHO84	YML123C	-8.7704	1.60E-36
			OPT2	YPR194C	-2.2986	4.49E-03
2Nv335	Up	10	YBR085C-A	YBR085C-A	1.7864	3.88E-02
			HXT6	YDR343C	3.1740	1.33E-07
			SIT1	YEL065W	1.7962	3.87E-02
			SPI1	YER150W	2.9213	2.68E-06
			BTN2	YGR142W	1.8232	3.69E-02
			YHR138C	YHR138C	1.8996	2.03E-02
			YJL133C-A	YJL133C-A	2.5168	6.27E-04
			GTO3	YMR251W	2.8562	4.54E-04
			HOR7	YMR251W-A	1.8922	2.03E-02
			FRE5	YOR384W	2.3913	1.20E-03
	Down	12	PHO5	YBR093C	-3.7294	5.80E-10
			PHO89	YBR296C	-4.6118	5.26E-14
			PHM6	YDR281C	-3.3973	2.68E-06
			ATO3	YDR384C	-1.9213	2.03E-02
			MIG2	YGL209W	-2.4946	9.32E-04
			SPL2	YHR136C	-4.2957	8.21E-09
			HMS2	YJR147W	-1.9988	1.56E-02
			CWP1	YKL096W	-1.8588	2.57E-02
			PTR2	YKR093W	-2.5314	2.27E-04
			PHO84	YML123C	-6.1548	5.77E-23
			AQR1	YNL065W	-2.0554	7.75E-03
			ARG1	YOL058W	-2.2693	1.17E-03
2Nv336	Up	4	HXT6	YDR343C	3.5879	7.73E-10
			SPI1	YER150W	1.7857	3.83E-02
			BTN2	YGR142W	1.9241	1.34E-02
			FBP1	YLR377C	2.1522	2.24E-03
	Down	25	PHO11	YAR071W	-3.7218	1.53E-03
			HSP26	YBR072W	-2.0355	9.99E-03
			PHO5	YBR093C	-4.0691	6.13E-12
			PHO89	YBR296C	-5.3176	1.06E-17
			PHM6	YDR281C	-4.4040	7.73E-10
			ARO10	YDR380W	-2.8562	1.53E-03
			POX1	YGL205W	-1.8463	3.90E-02
			MIG2	YGL209W	-1.9989	2.42E-02
			ZRT1	YGL255W	-2.0160	5.96E-03
			ECL1	YGR146C	-2.3450	1.53E-03
			SPL2	YHR136C	-5.4629	5.31E-12

			ARO9	YHR137W	-2.0378	1.88E-02
			PHO12	YHR215W	-3.4654	1.53E-03
			DCG1	YIR030C	-2.0372	4.96E-02
			ARG3	YJL088W	-2.3020	4.51E-03
			NCA3	YJL116C	-1.8437	4.63E-02
			PTR2	YKR093W	-2.0993	5.01E-03
			AQY2	YLL052C	-2.7058	1.19E-04
			YLL053C	YLL053C	-2.2681	4.39E-03
			PHO84	YML123C	-8.0761	4.36E-33
			HXT2	YMR011W	-2.9188	1.44E-06
			ARG1	YOL058W	-3.1130	3.06E-07
			SPS4	YOR313C	-3.0066	2.96E-05
			SFG1	YOR315W	-1.7642	4.63E-02
			GRE1	YPL223C	-2.2932	7.59E-03
2Nv337	Up	16	GPM2	YDL021W	1.8083	3.65E-02
			HXT6	YDR343C	3.4741	5.95E-09
			YDR461C-A	YDR461C-A	1.8473	3.01E-02
			SIT1	YEL065W	2.3949	3.37E-04
			SSA4	YER103W	2.2897	7.77E-04
			YFR032C-B	YFR032C-B	1.7722	3.25E-02
			GPG1	YGL121C	1.8888	2.05E-02
			BTN2	YGR142W	2.5903	8.14E-05
			TIS11	YLR136C	1.8021	3.77E-02
			COX7	YMR256C	1.7512	3.77E-02
			MFA2	YNL145W	2.1312	5.64E-03
			FIT2	YOR382W	2.5418	1.57E-04
			FIT3	YOR383C	2.2783	8.88E-04
			FRE5	YOR384W	1.8980	3.25E-02
			MF(ALPHA)1	YPL187W	2.0162	2.36E-02
			HSP82	YPL240C	2.3336	4.71E-04
	Down	17	PHO5	YBR093C	-3.2236	1.23E-07
			PHO89	YBR296C	-5.8873	8.33E-21
			ADY2	YCR010C	-2.3810	3.63E-03
			GIT1	YCR098C	-3.5541	4.34E-07
			PHM6	YDR281C	-2.8674	1.57E-04
			GLC3	YEL011W	-1.7374	3.77E-02
			FMP48	YGR052W	-2.0233	2.42E-02
			ECL1	YGR146C	-2.3038	1.72E-03
			MGA1	YGR249W	-2.0423	8.31E-03
			SPL2	YHR136C	-4.1623	2.48E-08
			CWP1	YKL096W	-2.3463	4.71E-04

			PTR2	YKR093W	-2.2756	1.20E-03
			MMP1	YLL061W	-2.0739	5.08E-03
			PHO84	YML123C	-6.9649	1.60E-27
			HXT2	YMR011W	-2.2230	1.11E-03
			ADH2	YMR303C	-1.9893	7.01E-03
			TPO4	YOR273C	-1.9876	7.01E-03

Table 12-1 Significantly differentially expressed genes in evolved clones relative to the 2N ancestor

Up/Down	Number of Genes	Gene name	Systematic Name	log_2fold Change	Adjusted P-value
		ATG8	YBL078C	2.0002	4.67E-03
		YBR056W-A	YBR056W-A	1.7973	1.76E-02
		YBR085C-A	YBR085C-A	1.6987	2.12E-02
		RTC2	YBR147W	1.7115	4.01E-02
		YBR285W	YBR285W	2.7896	1.12E-05
		YCL042W	YCL042W	1.7232	2.56E-02
		HSP30	YCR021C	2.4241	2.04E-04
		STF1	YDL130W-A	1.5889	3.98E-02
		RTN2	YDL204W	2.0071	4.17E-03
		Up	92	3.0676	1.11E-05
		FMP16	YDR070C	2.3601	9.06E-04
		YDR133C	YDR133C	1.9563	4.17E-03
		HSP42	YDR171W	2.4271	1.22E-04
		YDR381C-A	YDR381C-A	1.7494	2.17E-02
		Unknown	YDR461C-A	1.9986	4.77E-03
		GLC3	YEL011W	1.5824	3.95E-02
		CYC7	YEL039C	2.9930	1.27E-05
		DSF1	YEL070W	2.7339	3.08E-05
		PIC2	YER053C	1.9784	4.17E-03
		GIP2	YER054C	2.4606	1.22E-04
		RGI1	YER067W	2.6995	1.27E-05
		YER079W	YER079W	1.6206	4.04E-02
		YER121W	YER121W	2.1332	5.43E-03
		SPI1	YER150W	2.8403	6.71E-06
		HSP12	YFL014W	2.1438	1.21E-03
		GSY1	YFR015C	1.9461	4.36E-03
		YFR032C-B	YFR032C-B	1.8837	6.65E-03
		PNC1	YGL037C	1.5631	4.18E-02
		STF2	YGR008C	2.2252	5.30E-04
		YGR067C	YGR067C	1.8131	1.50E-02

Up/Down	Number of Genes	Gene name	Systematic Name	log_2fold Change	Adjusted P-value
		CTT1	YGR088W	2.4277	1.96E-04
		SOL4	YGR248W	3.2609	3.42E-07
		AIM17	YHL021C	1.8520	8.54E-03
		RTC3	YHR087W	2.6352	3.59E-05
		YHR097C	YHR097C	1.6580	2.79E-02
		CRG1	YHR209W	1.6886	2.77E-02
		RGI2	YIL057C	2.3974	1.96E-04
		XBP1	YIL101C	1.6790	2.67E-02
		COX5b	YIL111W	1.5948	4.01E-02
		PRM5	YIL117C	2.8735	1.35E-05
		OM45	YIL136W	1.8909	6.10E-03
		YIR016W	YIR016W	1.6973	2.12E-02
		YJL144W	YJL144W	1.9419	1.68E-02
		FMP33	YJL161W	2.2391	2.35E-03
		TPK1	YJL164C	1.9892	4.03E-03
		YJR115W	YJR115W	1.9910	7.48E-03
		GPX1	YKL026C	1.6526	3.14E-02
		PHD1	YKL043W	1.7811	1.45E-02
		YKL091C	YKL091C	1.6409	3.04E-02
		PIR3	YKL163W	2.7380	1.80E-05
		GLG1	YKR058W	1.6647	2.67E-02
		SRL3	YKR091W	2.3271	3.91E-04
		YLR149C	YLR149C	1.5863	4.18E-02
		YLR177W	YLR177W	1.6489	2.79E-02
		TFS1	YLR178C	1.6132	3.88E-02
		SYM1	YLR251W	1.9133	1.04E-02
		YLR252W	YLR252W	2.1910	1.44E-03
		GSY2	YLR258W	1.9153	5.29E-03
		YLR285C-A	YLR285C-A	2.2164	4.36E-03
		TMA10	YLR327C	2.5984	2.98E-05
		TSL1	YML100W	2.0043	3.22E-03
		MSC1	YML128C	2.5623	3.76E-05
		YET2	YMR040W	2.1194	2.38E-03
		ISF1	YMR081C	1.7081	1.99E-02
		YMR103C	YMR103C	1.6467	2.96E-02
		PGM2	YMR105C	2.2926	3.32E-04
		SPG4	YMR107W	1.8027	3.04E-02
		CMC4	YMR194C-B	1.8124	4.10E-02
		GAD1	YMR250W	1.8971	6.65E-03

Up/Down	Number of Genes	Gene name	Systematic Name	log_2fold Change	Adjusted P-value
		HOR7	YMR251W-A	2.3176	2.68E-04
		NCE103	YNL036W	2.2358	5.15E-04
		YPT53	YNL093W	1.7562	2.12E-02
		YNL194C	YNL194C	3.3111	3.42E-07
		YNL195C	YNL195C	2.6856	9.13E-05
		YNR034W-A	YNR034W-A	3.4232	7.76E-08
		YNR073C	YNR073C	2.3531	4.03E-03
		DDR2	YOL052C-A	3.0185	2.61E-06
		PHM7	YOL084W	1.6791	3.76E-02
		ZEO1	YOL109W	1.7542	1.52E-02
		YOR019W	YOR019W	2.1473	2.62E-03
		BAG7	YOR134W	1.9746	5.67E-03
		PNS1	YOR161C	1.6393	3.65E-02
		DCS2	YOR173W	2.5691	4.88E-05
		YOR199W	YOR199W	5.3000	5.15E-04
		RNY1	YPL123C	1.6170	3.97E-02
		PXA1	YPL147W	1.8011	1.54E-02
		UIP4	YPL186C	2.3765	2.04E-04
		USV1	YPL230W	1.8669	8.54E-03
		Unknown	YPR145C-A	2.4630	9.36E-03
		SUE1	YPR151C	1.8986	1.99E-02
		GPH1	YPR160W	2.0379	2.43E-03
		GDB1	YPR184W	1.7245	1.92E-02
Down	25	PHO5	YBR093C	-1.5785	4.25E-02
		PHO89	YBR296C	-2.0116	4.17E-03
		RSA4	YCR072C	-1.9151	5.88E-03
		BSC1	YDL037C	-1.8019	1.50E-02
		PHM6	YDR281C	-1.9496	1.99E-02
		SAM2	YDR502C	-1.6106	3.36E-02
		RRT5	YFR032C	-1.9724	1.80E-02
		MIG2	YGL209W	-3.4627	5.89E-07
		YGR079W	YGR079W	-2.4300	1.22E-04
		AIR1	YIL079C	-1.8138	1.43E-02
		UTP18	YJL069C	-1.5670	4.31E-02
		CDC6	YJL194W	-1.7031	2.75E-02
		YKL068W-A	YKL068W-A	-2.1585	1.22E-02
		YLR063W	YLR063W	-1.7181	3.04E-02
		IFH1	YLR223C	-1.6753	2.45E-02
		YML018C	YML018C	-1.5951	4.04E-02

Up/Down	Number of Genes	Gene name	Systematic Name	log_2fold Change	Adjusted P-value
		YOX1	YML027W	-1.7952	1.52E-02
		PHO84	YML123C	-2.0633	2.43E-03
		AQR1	YNL065W	-2.9480	3.08E-06
		TRF5	YNL299W	-1.8417	9.90E-03
		SPS4	YOR313C	-1.8420	2.42E-02
		SFG1	YOR315W	-2.2822	5.15E-04
		VTS1	YOR359W	-1.7630	1.62E-02
		SET6	YPL165C	-4.3204	2.89E-02
		SAM3	YPL274W	-1.6175	3.51E-02

Table 12-2 Significantly differentially expressed genes in the tetraploid ancestral strain relative to the 2N ancestral strain

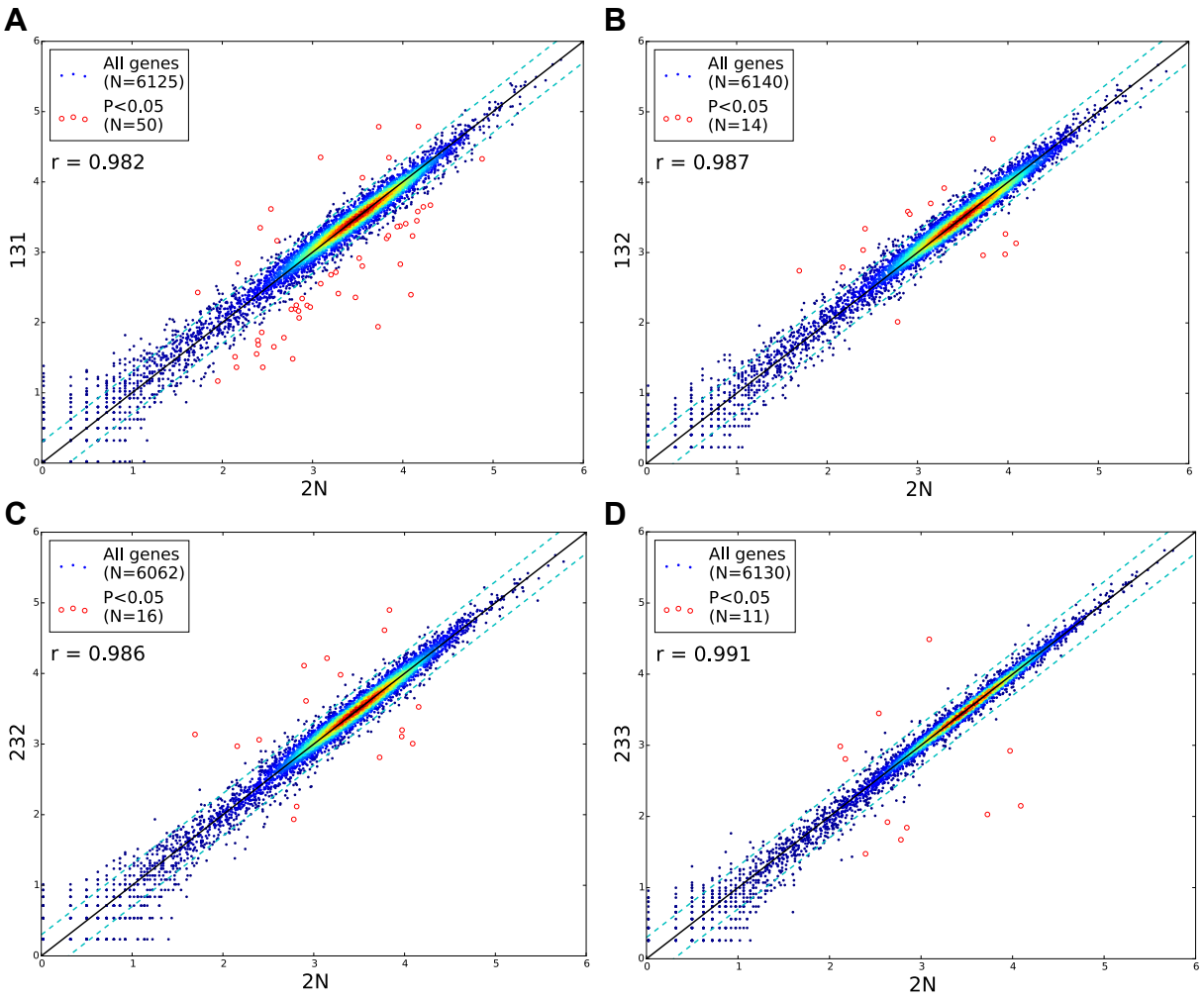


Figure 12-1 Comparative expression scatterplots for 1Ne and 2Ne clones Normalized $\log_{10}$ gene expression in the evolved clones (Y-axis) A) 131, B) 132, C) 232, and D) 233 compared to the diploid ancestor (X-axis). The gene expression values for each gene are the normalized read counts calculated by DESeq. Significantly differentially expressed genes (Adjusted P<0.05) are denoted by red circles. The density of genes is indicated by color from blue dots (low density) to red dots (high density). The dashed cyan lines indicate 2-fold differential expression.

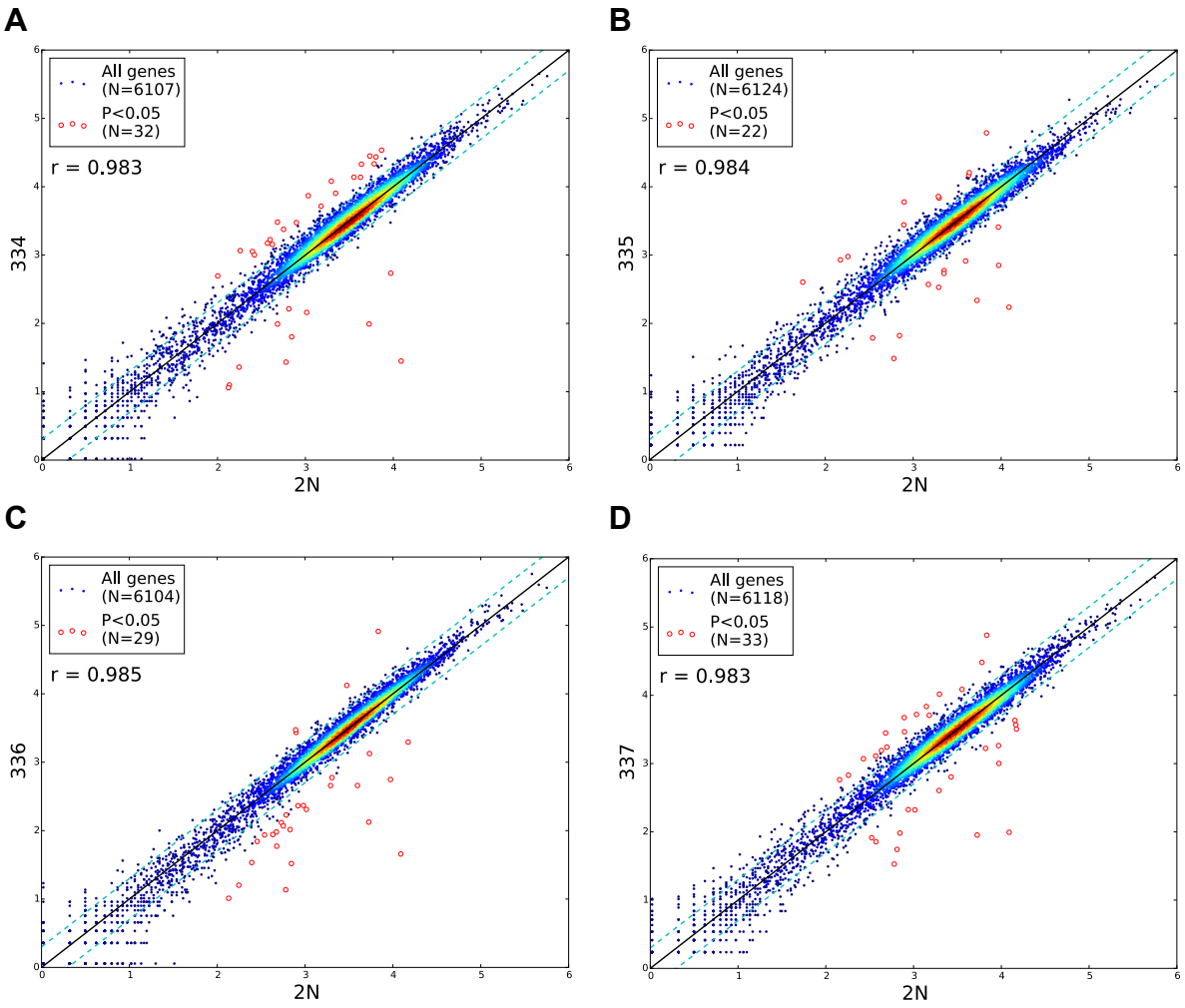


Figure 12-2 Comparative expression scatter plots of the 4Ne clones Normalized $\log_{10}$ gene expression in the evolved clones (Y-axis) A) 334, B) 335, C) 336, and D) 337 compared to the diploid ancestor (X-axis). The gene expression values for each gene are the normalized read counts calculated by DESeq. Significantly differentially expressed genes (Adjusted P<0.05) are denoted by red circles. The density of genes is indicated by color from blue dots (low density) to red dots (high density). The dashed cyan lines indicate 2-fold differential expression.

13 GENE ONTOLOGY ENRICHMENT ANALYSIS

Strains	GOID	GO_term	Cluster frequency	Background frequency	P-value	Gene(s) annotated to the term
6006v131	15749	monosaccharide transport	2 out of 11 genes, 18.2%	24 out of 5772 background genes, 0.4%	0.04276	HXT4/YHR092C:HXT2/YMR011W
6006v131	8645	hexose transport	2 out of 11 genes, 18.2%	24 out of 5772 background genes, 0.4%	0.04276	HXT4/YHR092C:HXT2/YMR011W
6006v132	NA	NA	NA	NA	NA	NA
6006v232	15749	monosaccharide transport	3 out of 11 genes, 27.3%	24 out of 5772 background genes, 0.4%	0.00049	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C
6006v232	8645	hexose transport	3 out of 11 genes, 27.3%	24 out of 5772 background genes, 0.4%	0.00049	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C
6006v232	8643	carbohydrate transport	3 out of 11 genes, 27.3%	33 out of 5772 background genes, 0.6%	0.00131	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C
6006v232	55085	transmembrane transport	5 out of 11 genes, 45.5%	232 out of 5772 background genes, 4.0%	0.00183	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:SIT1/YEL065W:SSA4/YER103W
6006v232	71702	organic substance transport	6 out of 11 genes, 54.5%	645 out of 5772 background genes, 11.2%	0.02573	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:SSA4/YER103W:BTN2/YGR142W:SFC1/YJR095W
6006v232	44765	single-organism transport	7 out of 11 genes, 63.6%	1021 out of 5772 background genes, 17.7%	0.04338	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:SIT1/YEL065W:SSA4/YER103W:BTN2/YGR142W:SFC1/YJR095W
6006v233	44699	single-organism process	2 out of 3 genes, 66.7%	3437 out of 5772 background genes, 59.5%	0	MIG2/YGL209W:HXT4/YHR092C
6006v334	15891	siderophore transport	3 out of 23 genes, 13.0%	8 out of 5772 background genes, 0.1%	0.00027	SIT1/YEL065W:FIT2/YOR382W:FIT3/YOR383C
6006v334	15688	iron chelate transport	3 out of 23 genes, 13.0%	9 out of 5772 background genes, 0.2%	0.0004	SIT1/YEL065W:FIT2/YOR382W:FIT3/YOR383C

6006v334	1901678	iron coordination entity transport	3 out of 23 genes, 13.0%	13 out of 5772 background genes, 0.2%	0.00134	SIT1/YEL065W:FIT2/YOR382W:FIT3/YOR383C
6006v334	6879	cellular iron ion homeostasis	3 out of 23 genes, 13.0%	35 out of 5772 background genes, 0.6%	0.02897	SIT1/YEL065W:TIS11/YLR136C:FET3/YMR058W
6006v334	55072	iron ion homeostasis	3 out of 23 genes, 13.0%	37 out of 5772 background genes, 0.6%	0.03421	SIT1/YEL065W:TIS11/YLR136C:FET3/YMR058W
6006v335	NA	NA	NA	NA	NA	NA
6006v336	15749	monosaccharide transport	5 out of 9 genes, 55.6%	24 out of 5772 background genes, 0.4%	2.78E-09	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:HXT4/YHR092C:GAL2/YLR081W
6006v336	8645	hexose transport	5 out of 9 genes, 55.6%	24 out of 5772 background genes, 0.4%	2.78E-09	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:HXT4/YHR092C:GAL2/YLR081W
6006v336	8643	carbohydrate transport	5 out of 9 genes, 55.6%	33 out of 5772 background genes, 0.6%	1.54E-08	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:HXT4/YHR092C:GAL2/YLR081W
6006v336	71702	organic substance transport	6 out of 9 genes, 66.7%	645 out of 5772 background genes, 11.2%	0.00333	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:BTN2/YGR142W:HXT4/YHR092C:GAL2/YLR081W
6006v336	51179	localization	7 out of 9 genes, 77.8%	1329 out of 5772 background genes, 23.0%	0.02185	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:BTN2/YGR142W:HXT4/YHR092C:HSP104/YLL026W:GAL2/YLR081W
6006v336	44765	single-organism transport	6 out of 9 genes, 66.7%	1021 out of 5772 background genes, 17.7%	0.04366	HXT7/YDR342C:HXT6/YDR343C:HXT3/YDR345C:BTN2/YGR142W:HXT4/YHR092C:GAL2/YLR081W
6006v337	15891	siderophore transport	3 out of 22 genes, 13.6%	8 out of 5772 background genes, 0.1%	0.00032	SIT1/YEL065W:FIT2/YOR382W:FIT3/YOR383C
6006v337	15688	iron chelate transport	3 out of 22 genes, 13.6%	9 out of 5772 background genes, 0.2%	0.00047	SIT1/YEL065W:FIT2/YOR382W:FIT3/YOR383C
6006v337	1901678	iron coordination entity transport	3 out of 22 genes, 13.6%	13 out of 5772 background genes, 0.2%	0.0016	SIT1/YEL065W:FIT2/YOR382W:FIT3/YOR383C
6006v337	15749	monosaccharide transport	3 out of 22 genes, 13.6%	24 out of 5772 background genes, 0.4%	0.01099	HXT6/YDR343C:GAL2/YLR081W:SKS1/YPL026C

6006v337	8645	hexose transport	3 out of 22 genes, 13.6%	24 out of 5772 background genes, 0.4%	0.01099	HXT6/YDR343C:GAL2/YLR081W:SKS1/YPL026C
6006v337	8643	carbohydrate transport	3 out of 22 genes, 13.6%	33 out of 5772 background genes, 0.6%	0.02898	HXT6/YDR343C:GAL2/YLR081W:SKS1/YPL026C

Table 13-1 GO terms enriched in genes up-regulated in the evolved clones relative to the 2N ancestor

Strains	GOID	GO_term	Cluster frequency	Background frequency	P-value	Gene(s) annotated to the term
6006v131	6820	anion transport	7 out of 43 genes, 16.3%	136 out of 5772 background genes, 2.4%	0.01113	PHO89/YBR296C:ADY2/YCR010C:GIT1/YCR098C:SFC1/YJR095W:JEN1/YKL217W:PHO84/YML123C:OPT2/YPR194C
6006v131	6112	energy reserve metabolic process	4 out of 43 genes, 9.3%	34 out of 5772 background genes, 0.6%	0.02138	GLC3/YEL011W:RGI1/YER067W:RGI2/YIL057C:GAC1/YOR178C
6006v131	97080	plasma membrane selenite transport	2 out of 43 genes, 4.7%	3 out of 5772 background genes, 0.1%	0.03286	JEN1/YKL217W:PHO84/YML123C
6006v132	6817	phosphate ion transport	2 out of 5 genes, 40.0%	10 out of 5772 background genes, 0.2%	0.00054	PHO89/YBR296C:PHO84/YML123C
6006v132	15698	inorganic anion transport	2 out of 5 genes, 40.0%	24 out of 5772 background genes, 0.4%	0.00329	PHO89/YBR296C:PHO84/YML123C
6006v132	98656	anion transmembrane transport	2 out of 5 genes, 40.0%	31 out of 5772 background genes, 0.5%	0.00553	PHO89/YBR296C:PHO84/YML123C
6006v132	34220	ion transmembrane transport	2 out of 5 genes, 40.0%	55 out of 5772 background genes, 1.0%	0.01751	PHO89/YBR296C:PHO84/YML123C
6006v232	NA	NA	NA	NA	NA	NA
6006v233	6817	phosphate ion transport	2 out of 9 genes, 22.2%	10 out of 5772 background genes, 0.2%	0.00232	PHO89/YBR296C:PHO84/YML123C
6006v233	15698	inorganic anion transport	2 out of 9 genes, 22.2%	24 out of 5772 background genes, 0.4%	0.01407	PHO89/YBR296C:PHO84/YML123C

6006v233	98656	anion transmembrane transport	2 out of 9 genes, 22.2%	31 out of 5772 background genes, 0.5%	0.02356	PHO89/YBR296C:PHO84/YML123C
6006v334	6817	phosphate ion transport	2 out of 13 genes, 15.4%	10 out of 5772 background genes, 0.2%	0.01377	PHO89/YBR296C:PHO84/YML123C
6006v335	6817	phosphate ion transport	2 out of 13 genes, 15.4%	10 out of 5772 background genes, 0.2%	0.00626	PHO89/YBR296C:PHO84/YML123C
6006v335	15698	inorganic anion transport	2 out of 13 genes, 15.4%	24 out of 5772 background genes, 0.4%	0.03771	PHO89/YBR296C:PHO84/YML123C
6006v335	55085	transmembrane transport	4 out of 13 genes, 30.8%	232 out of 5772 background genes, 4.0%	0.04095	PHO89/YBR296C:ATO3/YDR384C:PHO84/YML123C:AQR1/YNL065W
6006v335	6811	ion transport	4 out of 13 genes, 30.8%	235 out of 5772 background genes, 4.1%	0.04296	PHO89/YBR296C:ATO3/YDR384C:PHO84/YML123C:AQR1/YNL065W
6006v336	6817	phosphate ion transport	2 out of 23 genes, 8.7%	10 out of 5772 background genes, 0.2%	0.04693	PHO89/YBR296C:PHO84/YML123C
6006v337	15696	ammonium transport	3 out of 22 genes, 13.6%	16 out of 5772 background genes, 0.3%	0.00242	ADY2/YCR010C:TPO4/YOR273C:SAM3/YPL274W
6006v337	6556	S-adenosylmethionine biosynthetic process	2 out of 22 genes, 9.1%	3 out of 5772 background genes, 0.1%	0.00386	SAM2/YDR502C:SAM1/YLR180W
6006v337	15848	spermidine transport	2 out of 22 genes, 9.1%	4 out of 5772 background genes, 0.1%	0.0077	TPO4/YOR273C:SAM3/YPL274W
6006v337	46500	S-adenosylmethionine metabolic process	2 out of 22 genes, 9.1%	4 out of 5772 background genes, 0.1%	0.0077	SAM2/YDR502C:SAM1/YLR180W
6006v337	6811	ion transport	6 out of 22 genes, 27.3%	235 out of 5772 background genes, 4.1%	0.01712	PHO89/YBR296C:ADY2/YCR010C:GIT1/YCR098C:PHO84/YML123C:TPO4/YOR273C:SAM3/YPL274W
6006v337	15695	organic cation transport	2 out of 22 genes, 9.1%	7 out of 5772 background genes, 0.1%	0.02678	TPO4/YOR273C:SAM3/YPL274W

6006v337	72337	modified amino acid transport	2 out of 22 genes, 9.1%	8 out of 5772 background genes, 0.1%	0.03562	MMP1/YLL061W:SAM3/YPL274W

Table 13-2 GO terms enriched in the genes down-regulated in the evolved clones relative to the 2N ancestor

Strains	GOID	GO_term	Cluster frequency	Background frequency	P-value	Gene(s) annotated to the term
6006v6040	6112	energy reserve metabolic process	10 out of 105 genes, 9.5%	34 out of 5772 background genes, 0.6%	8.37E-08	GLC3/YEL011W:GIP2/YER054C:RGI1/YER067W:GSY1/YFR015C:RGI2/YIL057C:GLG1/YKR058W:GSY2/YLR258W:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	5977	glycogen metabolic process	8 out of 105 genes, 7.6%	32 out of 5772 background genes, 0.6%	2.39E-05	GLC3/YEL011W:GIP2/YER054C:GSY1/YFR015C:GLG1/YKR058W:GSY2/YLR258W:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	15980	energy derivation by oxidation of organic compounds	13 out of 105 genes, 12.4%	126 out of 5772 background genes, 2.2%	0.00012	GLC3/YEL011W:CYC7/YEL039C:GIP2/YER054C:RGI1/YER067W:GSY1/YFR015C:RGI2/YIL057C:COX5B/YIL111W:GLG1/YKR058W:GSY2/YLR258W:ISF1/YMR081C:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	6091	generation of precursor metabolites and energy	13 out of 105 genes, 12.4%	149 out of 5772 background genes, 2.6%	0.00086	GLC3/YEL011W:CYC7/YEL039C:GIP2/YER054C:RGI1/YER067W:GSY1/YFR015C:RGI2/YIL057C:COX5B/YIL111W:GLG1/YKR058W:GSY2/YLR258W:ISF1/YMR081C:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	44042	glucan metabolic process	8 out of 105 genes, 7.6%	51 out of 5772 background genes, 0.9%	0.00109	GLC3/YEL011W:GIP2/YER054C:GSY1/YFR015C:GLG1/YKR058W:GSY2/YLR258W:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	6073	cellular glucan metabolic process	8 out of 105 genes, 7.6%	51 out of 5772 background genes, 0.9%	0.00109	GLC3/YEL011W:GIP2/YER054C:GSY1/YFR015C:GLG1/YKR058W:GSY2/YLR258W:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	55114	oxidation-reduction process	13 out of 105 genes, 12.4%	153 out of 5772 background genes, 2.7%	0.00116	GLC3/YEL011W:CYC7/YEL039C:GIP2/YER054C:RGI1/YER067W:GSY1/YFR015C:RGI2/YIL057C:COX5B/YIL111W:GLG1/YKR058W:GSY2/YLR258W:ISF1/YMR081C:PGM2/YM

						R105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	5978	glycogen biosynthetic process	5 out of 105 genes, 4.8%	18 out of 5772 background genes, 0.3%	0.00459	GLC3/YEL011W:GSY1/YFR015C:GLG1/YKR058W:GSY2/YLR258W:PGM2/YMR105C
6006v6040	6979	response to oxidative stress	10 out of 105 genes, 9.5%	102 out of 5772 background genes, 1.8%	0.00487	HSP30/YCR021C:HSP12/YFL014W:NQM1/YGR043C:CTT1/YGR088W:XBP1/YIL101C:GPX1/YKL026C:HSP104/YLL026W:GAD1/YMR250W:NCE103/YNL036W:DDR2/YOL052C-A
6006v6040	44264	cellular polysaccharide metabolic process	8 out of 105 genes, 7.6%	63 out of 5772 background genes, 1.1%	0.00554	GLC3/YEL011W:GIP2/YER054C:GSY1/YFR015C:GLG1/YKR058W:GSY2/YLR258W:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	5976	polysaccharide metabolic process	8 out of 105 genes, 7.6%	69 out of 5772 background genes, 1.2%	0.01093	GLC3/YEL011W:GIP2/YER054C:GSY1/YFR015C:GLG1/YKR058W:GSY2/YLR258W:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	5991	trehalose metabolic process	4 out of 105 genes, 3.8%	11 out of 5772 background genes, 0.2%	0.01104	TPS2/YDR074W:HSP104/YLL026W:TSL1/YML100W:PGM2/YMR105C
6006v6040	44262	cellular carbohydrate metabolic process	11 out of 105 genes, 10.5%	138 out of 5772 background genes, 2.4%	0.01277	TPS2/YDR074W:GLC3/YEL011W:GIP2/YER054C:GSY1/YFR015C:GLG1/YKR058W:HSP104/YLL026W:GSY2/YLR258W:TSL1/YML100W:PGM2/YMR105C:GPH1/YPR160W:GDB1/YPR184W
6006v6040	34637	cellular carbohydrate biosynthetic process	7 out of 105 genes, 6.7%	55 out of 5772 background genes, 1.0%	0.01907	TPS2/YDR074W:GLC3/YEL011W:GSY1/YFR015C:GLG1/YKR058W:GSY2/YLR258W:TSL1/YML100W:PGM2/YMR105C
6006v6040	34599	cellular response to oxidative stress	9 out of 105 genes, 8.6%	96 out of 5772 background genes, 1.7%	0.0192	HSP30/YCR021C:HSP12/YFL014W:NQM1/YGR043C:XBP1/YIL101C:GPX1/YKL026C:HSP104/YLL026W:GAD1/YMR250W:NCE103/YNL036W:DDR2/YOL052C-A

Table 13-3 GO terms enriched in the genes up-regulated in the 4N ancestor relative to the 2N ancestor

Strains	GOID	GO_term	Cluster frequency	Background frequency	P-value	Gene(s) annotated to the term
6006v6040	15847	putrescine transport	2 out of 31 genes, 6.5%	4 out of 5772 background genes, 0.1%	0.03211	TPO1/YLL028W:SAM3/ YPL274W
6006v6040	15848	spermidine transport	2 out of 31 genes, 6.5%	4 out of 5772 background genes, 0.1%	0.03211	TPO1/YLL028W:SAM3/ YPL274W

Table 13-4 GO terms enriched in the genes down-regulated in the 4N ancestor relative to the 2N ancestor

14 PRIMERS USED IN THESE STUDIES

Primer Name	Sequence	Assay
PR001F_HXT1	GTTTGTGTTTCGCCTGGGCCTT	qPCR
PR001R_HXT1	TGGATGGTCAGGTGGGCATTTG	qPCR
PR002F_HXT2	CGCTACTAGCCGCGTTGAAAGT	qPCR
PR002R_HXT2	ATGCGGCGATTGGCTTTGCT	qPCR
PR003F_HXT3	TTTCCAAGCTGAGGCCGACCAA	qPCR
PR003R_HXT3	ACGAAACCACCGAAGGCAACCA	qPCR
PR004F_HXT4	TAAGGTCAGCGCAGACGATCCA	qPCR
PR004R_HXT4	TTCACCCCAGGAGGCATTACCA	qPCR
PR005F_HXT5	TGGTACCGCCGGTTACAACGAT	qPCR
PR005R_HXT5	TCGTCTTTGGGAGGGCCTTCAT	qPCR
PR006F_HXT67	ATGGGGTGCTGCATCCATGACT	qPCR
PR006R_HXT67	GCACCCTTGGAAGATGGTTGGT	qPCR
PR007F_MIG2	ACACTCCACGCTCTGTGCCAAA	qPCR
PR007R_MIG2	AGGGCTGGCAGTTTGATGCTGA	qPCR
PR008F_SUC2	TTCACACCCAACAAGGGCTGGA	qPCR
PR008R_SUC2	TGGCCCCAAAACAATGGCGT	qPCR
PHO89_F	AAGCCATGGTTTTGGCGGGT	qPCR
PHO89_R	AAAACAGCGGGGTCGTTGGT	qPCR
PR010F_SNF3	AGGGCATAGCACTGCGACGAAA	qPCR
PR010R_SNF3	TTTTGCCGCTGCAAGCTCGT	qPCR
PR016F_MTH1	TGAGATCAAGCGCCGCAACA	qPCR
PR016R_MTH1	ATCCGGCTGCCAATCCAATCCT	qPCR
PR013F_ACT1	ACGTCGCCTTGGACTTCGAACA	qPCR
PR013R_ACT1	TGGAACAAAGCTTCTGGGGCTC	qPCR
ZPS1_F	CCGTCATGGGTGTCTTTGAGCA	qPCR
ZPS1_R	AGCTGGAGCAGATTGACGGTGA	qPCR
NCE102_F	TGTTGGCCGTTGGTATCAGAGC	qPCR
NCE102_R	TGCAACAGCGGCTTGAGCTTGT	qPCR
YIL169C_F	AGCTGCTTTAGCTTTGGCCCTT	qPCR
YIL169C_R	ACCACAGAGGTGGATGAGCTGT	qPCR
SPS4_F	TGGTAGTGTCACGTCCCGTTCA	qPCR
SPS4_R	AAGGGACAACGCGCTCTACTGT	qPCR
SPI1_F	TAACGATACAGCCACGCCAGCA	qPCR
SPI1_R	CGAAAGTCGTTGGTTCTGGGCA	qPCR
PR009F_PHO84	ATGGAGAGGTGCCATCATGGGT	qPCR
PR009R_PHO84	TCGCCCTTGTAAGCAGCAACCA	qPCR
SPL1_F	TGTCTCCACGCGGCAAAATGGA	qPCR
SPL1_R	AGGATTGAGCCTCCTGCACTCT	qPCR

Primer Name	Sequence	Assay
AS001_F_Flo8	GACACGGTGAGTTGACGTTAG	Gene Amp
AS001_R_Flo8	CGTAACTCCATTCTCCTAGCTTTT	Gene Amp
AS046: RPS12-Chip F1	TTGCTTGCGGGTAAGGTATT	Rap1 Chip-qPCR
AS046: RPS12-Chip R1	TATTTCGGGGTTTGTTTCGG	Rap1 Chip-qPCR
AS047: RPL30-Chip F1	TCATTACTTTGTAGGCGGGA	Rap1 Chip-qPCR
AS047: RPL30-Chip R1	TTTAAATGCGGCCCTAGCTG	Rap1 Chip-qPCR
AS048: RNR2-Chip F1	TAGCACCGTACCATACCCTT	Rap1 Chip-qPCR
AS048: RNR2-Chip R1	CGAAAAAGGGGAAAGTGATGA	Rap1 Chip-qPCR
AS049: FAS1-Chip F1	AAACGACGGCCAAAAACTTC	Rap1 Chip-qPCR
AS049: FAS1-Chip R1	TACTGTCAGACCAAAAGCGT	Rap1 Chip-qPCR
AS050: PHO5-Chip F1	TCTCTTCGAAAACAGGGACC	Pho4 Chip-qPCR
AS050: PHO5-Chip R1	TTTCGCATAGAACGCAACTG	Pho4 Chip-qPCR
AS051: PHO8-Chip F1	TGGCCTTTTTGATCGCATTT	Pho4 Chip-qPCR
AS051: PHO8-Chip R1	AGTCATGTCGTACAACGGAA	Pho4 Chip-qPCR
AS052: VTC4-Chip F1	GTCAGTAGCTCTCCGTCAAA	Pho4 Chip-qPCR
AS052: VTC4-Chip R1	GAAGCGTTGCTAATATTCCGA	Pho4 Chip-qPCR
AS053: PHO84-Chip F1	GTGCTGGAAATAACACGTCC	Pho4 Chip-qPCR
AS053: PHO84-Chip R1	GCACGTTGGTGCTGTTATAG	Pho4 Chip-qPCR
AS055-MTH1seq-F1	CGAAACCACAAGCAGCAATA	Sanger
AS055-MTH1seq-R1	ACCATCGGGAAGGTTTCTTT	Sanger
AS055-MTH1seq-F2	AGTGAGCCTGGATGAAGCAT	Sanger
AS056-SNF3seq-F1	ATGCCTTTGTTGGCATAGAA	Sanger
AS056-SNF3seq-R1	ATAATGCACGTCCGCTTAAT	Sanger
AS056-SNF3seq-F2	GGTGCTGGAGGAATCACATT	Sanger
AS056-SNF3seq-F3	ATTGCCCTTCAAGCATTTCA	Sanger
AS056-SNF3seq-F4	TGACCGTTTATGAAACGAAGG	Sanger
AS056-SNF3seq-F5	CAAGCGAAGATTACACAGAAGATG	Sanger
AS101_MIG1_F	GAAGGTTGTGGGCTCTCCAA	qPCR
AS101_MIG1_R	TTGGAGAGCCCACAACCTTC	qPCR
AS102_HXK1_F	CTTCTCGTACCCAGCTTCCC	qPCR
AS102_HXK1_R	ACATCGTGGCCTTCGACATT	qPCR
AS103_GAL4_F	AAGAAAAACCGAAGTGCGCC	qPCR
AS103_GAL4_R	AGTCAGCGGAGACCTTTTGG	qPCR
AS104_GAL1_F	GGATCAGGCTGCCTCTGTTT	qPCR
AS104_GAL1_R	ACGGAGTAGCCTTCAACTGC	qPCR
AS105_STD1_F	GTCGAGGCCCAATGCAGATA	qPCR
AS105_STD1_R	ATGCGGTATGAGGCTTGGAG	qPCR
AS106_PFK27_F	TGTTATTAATGCGGGCGTCG	qPCR
AS106_PFK27_R	CCCTAATTTGCTTGCCGCAG	qPCR
AS107_IPT1_F	CACGATTGGCCCTAACGACT	qPCR
AS107_IPT1_R	AGATGAAGGGGAGGAGGACC	qPCR

Primer Name	Sequence	Assay
AS108_SIP4_F	TGTCGAGGAGAGGTCTTCCA	qPCR
AS108_SIP4_R	AGAATTTGCATTGGCGCTGG	qPCR
AS109_DUR3_F	TTTGGCCCATGCCCATGTAT	qPCR
AS109_DUR3_R	TACAAACGGCAAACGCACTG	qPCR
AS110_SNF3delta_F	CTATCCTCGGCAAATTGCAT	Gene amp
AS110_SNF3delta_R	ATAATGCACGTCCGCTTAAT	Gene amp
AS111_IPT1delta_F	TCCGGCCACAACATTTTTAT	Gene amp
AS111_IPT1delta_R	GTGGGATGTTCGTCGTTCTT	Gene amp
AS112_MOT3delta_F	GGCCTTGTAGCAATCAAAAA	Gene amp
AS112_MOT3delta_R	AGATGACGATTTCCCTCACG	Gene amp
AS113_IPT1_C219W_F	GGACCCCCGCTATATTTTGT	Allele flanking
AS113_IPT1_C219W_R	CATCCTGCACTTGACAGCTC	Allele flanking
AS113_IPT1_C219W_G_F	TTTGTAATCCTAGGGCAAAACTAAAG	Allele specific
AS113_IPT1_C219W_C_F	TTTGTAATCCTAGGGCAAAACTAAAC	Allele specific
AS115_DUR3delta_F	CGGTGTTGCATGAAGATACG	Gene amp
AS115_DUR3delta_R	TTTTTCCCCTTGACTTTCTTTTC	Gene amp
AS121_CYC7_qPCR_F	CGGGATTCAAACCAGGCTCT	qPCR
AS121_CYC7_qPCR_R2	CTTCTTCAACCCGGCAAACG	qPCR
AS122_CYC1_qPCR_F	CAAGGCCGGTTCTGCTAAGA	qPCR
AS122_CYC1_qPCR_R	TGGACCAACCTTATGTGGGC	qPCR
AS123_PGM2_qPCR_F	GCCAGCATGGTCTTCTGTCT	qPCR
AS123_PGM2_qPCR_R	CCTGGATTATGTGAGGCGGT	qPCR
As124_PGM1_qPCR_F	TTGAGGTTATCCGGCACAGG	qPCR
AS124_PGM1_qPCR_R	GAAGACGTCAGCTGTTTGGC	qPCR
AS125_GLG1_qPCR_F	CCGGACGTGTACGAGTCAAA	qPCR
AS125_GLG1_qPCR_R	TCTGCAACTGCACTCGTCTT	qPCR
AS126_GLG2_QPCR_F	CTGGTGCAGTTCGATCAGGT	qPCR
AS126_GLG2_qPCR_R2	GTTGTAGGTGAAGGGGAGGC	qPCR
AS127_TPK1_qPCR_F	ACCACCCCTGGTTCAAAGAA	qPCR
AS127_TPK1_qPCR_R	GGTGTCACCTTGTCCCTGTT	qPCR
AS128_TPK3_qPCR_F	GCCGGAAGTGGTCAGTACAA	qPCR
AS128_TPK3_qPCR_R	GGAGTGTATCCGGCAAGCAT	qPCR
AS129_CTT1_qPCR_F	AGACCAGACGGCCCTATCTT	qPCR
AS129_CTT1_qPCR_R	TTTGGCATGGACTACACGCT	qPCR
AS130_CTA1_qCPR_F	CAAACAATGACCGAACGCGA	qPCR

Primer Name	Sequence	Assay
AS130_CTA1_qPCR_R	AATTGCCCCTGAGGCCATAC	qPCR
AS135 RGT1 qPCR F	AAGGGGGTCTGCTCAAACTG	qPCR
AS135 RGT1 qPCR R	TGGTGCTCCTCGTATAGCCT	qPCR
AS137_ACT1_qPCR_F	ACGTCGCCTTGGACTTCGAACA	qPCR
AS137_ACT1_qPCR_R	TGGAACAAAGCTTCTGGGGCTC	qPCR

Table 14-1 Yeast Primer List

Primer Name	Sequence
AS117_MTH1_C321F_gblock	CGCTAGAAAAAGACATGAGCATATAACTGGTAGAAATGAAGCCGTCATGAATTTGTCGAAACCGGAATGGAGAAATATCATCGAAAATTACCTCTTAAATATAGCAGTAGAGGCACAATTCAGGTTTGATTTCAAACAAAGATGCTCCGAATATAAGAAATGGAAGTTACAACAGTCCAACTTAAAAAGACCGGACATGCCCCCACCAAGCATAATACCGCGGAAAAACAGCACAGAAAC
AS116_MTH1_C321F_pcore_F	AACCGGAATGGAGAAATATCATCGAAAATTACCTCTTAAATATAGCAGTAGAGGCACAATGAGCTCGTTTTCGACACTGG
AS116_MTH1_C321F_pcore_R	TTGGACTGTTGTAACTTCCATTTCTTATATTCGGAGCATCTTTGTTTGAAATCAAACCTGTCCTTACCATTAAGTTGATC
AS119_MTH1_gBLOCKamp_F	CGCTAGAAAAAGACATGAGCA
AS119_MTH1_gBLOCKamp_R	TTTTCCGCGGTATTATGCTT
AS131_MTH1-pRS_F	TGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCCGAGTCCATTTCTCCAGTGA
AS131_MTH1-pRS_R	TCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGTGGTTTGATCTTCGCTACCC
AS131_MTH1-pRS_F2	TGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATCCCCACCAAACCCAGTTTTCT
AS141_MTH1_C321F_gap_R1	AGCATCTTTGTTTGAAATCAAACCTGAATTGTGCCTC
AS142_MTH1_C321F_gap_F1	TAAATATAGCAGTAGAGGCACAATTCAGGTTTGA

Table 14-2 MTH1-C321F strain construction primers

Primer Name	Sequence
AS118_RGT1_S509 Stop_gblock	GCCACAGAGAAGTACAAAAAAACGCAGAAAAAGTTACGTATCTAAGAAGAC AAAACCAAAGAGAGATTCATCTATATCTATAACATCGAAAGATTCTGCTCAC CCAATGACCACTTCATGAACTATCGCGTATGGACAGATATCCGATGTAGAT CTAATAGACACCTACTATGAGTTCATACATGTAGGATTTCCGATCATACCTT TAAACAAAACGACCTTGACCAGTGACTTATTGTT
AS117_RGT1_S509 Stop_pcore_F	AGAGAGATTCATCTATATCTATAACATCGAAAGATTCTGCTCACCCAATGAC CACTTCATGAGCTCGTTTTCGACACTGG
AS117_RGT1_S509 Stop_pcore_R	ATGAACTCATAGTAGGTGTCTATTAGATCTACATCGGATATCTGTCCATAC GCGATAGTTTCCTTACCATTAAGTTGATC
AS120_RGT1_gBlo ckamp_F	CGCAGAAAAAGTTACGTATCTAAGAA
AS120_RGT1_gBlo ckamp_R	CACTGGTCAAGGTCGTTTTG
AS132_RGT1-pRS_F	TGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATC GCTGCTTTGTGCATTTTCCT
AS132_RGT1-pRS_R	TCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCA GGCTTGGCGACTCTCGAATAC
AS132_RGT1-pRS_F2	TGGGTACCGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTGATATC TTCTCTGGCTTTTTCGGT
AS143_RGT1-S509Stop_Gap_R1	ATATCTGTCCATACGCGATAGTTCATGAAGTGGTC
AS144_RGT1-S509Stop_Gap_F1	TCACCCAATGACCACTTCATGAACTATCGCG

Table 14-3 RGT1-S509stop strain construction primers

Primer Name	Sequence
399_IPT1_GBLOCK	ATGGCGGAGTGTAGTGAAGGAACAGCACCAAACACAATAGGAGACATGT GAAATCCTTTAGTGTTCAAATGAGTACCCAAATGGCTATCTACTCTTATCA AACCGGCAGCAAACCCTTCCTGGGTGTAATTAACATGTTCGGTGTCATCA ATACCGTACAGATGAGTAAACCATGGGGAGGCCATTGGAACAAGCAGAT GGGTAAGGACCCCCGCTATATTTTGTAATCCTAGGGCAAAACTAAACCAT TTTAAAGTCCCCGGTGGTTGGAAGACGTATAAGTATACAGCTGTTAAAAT TGGAGCTGTCAAGTGCAGGATGACATAGGAGAACCAAGCCAGTAAATCC TTGGTCTTAGTAAAGTTTGCATCCGACTGGTAAGCAAAGAAATGATCAAT GTTGAGTATAACAAATGTAAACAGGGGCAAGGCAAACGGTATGAAAATTG TCCGAAATCTCTTTTGGATCGGCCGCTGCTTGCGCTTGCAGGGGTAATG AAAA
397_IPT1_PCORE_F	GATGGGTAAGGACCCCCGCTATATTTTGTAATCCTAGGGCAAAACTAAAG GAGCTCGTTTTCGACACTGG
398_IPT1_Pcore_R	TTAACAGCTGTATACTTATACGTCTTCCAACCACCGGGGACTTTAAAATGT CCTTACCATTAAGTTGATC
400_IPT1_GBLOCK AMP_F	ATGGCGGAGTGTAGTGAAGG
401_IPT1_GBLOCK AMP_R	TTTTCATTACCCCTGCAAGC
AS138_IPT1_GBLO CKAMP_F2	AGGCCATTGGAACAAGCAGA
AS138_IPT1_GBLO CKAMP_R2	GTCATCCTGCACTTGACAGC

Table 14-4 IPT1-C219W strains construction primers

Primer Name	Sequence
419_MOT3_GBLOCK	GCATCTACCTCCTGGTTGGAAAATAAACACTATGCCGCAACCACGTC CTACGACAGCACCTAACCATCCCCCTGCGCCGGTGCCTTCTTCGAA CCCTGTGGCCTCGAACTTGGTTCCTGCCCCATCATCAGACCATAAAT ATATCCATCAATGCCAATTTTGTGAGAAGTCTTTCAAAAGAAAATCAT GGTTGAAAAGGCACCTATTGTCACACTCGCAACAAAGACATTTTCTA TGCCCTTGGTGCTTAAGCAGGCAGAAGAGAAAAGATAATCTTTTACA GCATATGTAACTCAAGCATACAAATTATTTATTAGACGAACTCAAGAA AAACAACATCATCTTTAACTACAACAATTCTTCCTCCTCTAATAATAAC AACGACAATAATAATAATAATAACAGCAATAGCGCTAGCGGCAGTGG CGGTGCCGGTGCCGCGGCGGCAGCAGCAACAGCTCCCGAAAATGA AGATGGAAACGGTTACG
417_MOT3_Pcore_F	TTGGTGCTTAAGCAGGCAGAAGAGAAAAGATAATCTTTTACAGCATA TGAGAGCTCGTTTTCGACACTGG
418_MOT3_Pcore_R	GATGTTGTTTTTCTTGAGTTCGTCTAATAAATAATTTGTATGCTTGAGT TTCCTTACCATTAAGTTGATC
420_MOT3_GBLOCKAMP_F	GCATCTACCTCCTGGTTGGA
421_MOT3_GBLOCKAMP_R	CGTAACCGTTTCCATCTTCA
AS139_MOT3_GBLOCKAMP_F2	TTTCTATGCCCTTGGTGCTT
AS139_MOT3_GBLOCKAMP_R2	CCGCCACTGCCGCTA

Table 14-5 MOT3-K394stop strain construction primers

Primer Name	Sequence	Assay
AS-FOXA2-F1	TACGTGTTCATGCCGTTCAT	qPCR
AS-FOXA2-R1	CGACTGGAGCAGCTACTATGC	qPCR
AS-FOXA2-F2	TGTTGCTCACGGAGGAGTAG	qPCR
AS-FOXA2-R2	TTAAAGTATGCTGGGAGCGG	qPCR
AS-REST-F1	GAGGCCACATAACTGCACTG	qPCR
AS-REST-R1	TGTCCTTACTCAAGTTCTCAGAAGA	qPCR
AS-REST-F2	CCACATAACTGCACTGATCACA	qPCR
AS-REST-R2	CATACAGGAGAACGCCCATA	qPCR
AS-GABPA-F1	TCAGCTCCTCTGCTTCTCTTTT	qPCR
AS-GABPA-R1	CTCAGCCGGCTCTGGAGT	qPCR
AS-GABPA-F2	TGGCTGGAGTATTTCAAAGGA	qPCR
AS-GABPA-R2	TCTTACCCGGAGAGACGCT	qPCR
AS-BACH1-F1	TGTTGTCGGGAAGTTCAGTG	qPCR
AS-BACH1-R1	GCTCTCGCTTCAGTCAGTCG	qPCR
AS-BACH1-F2	TCGGGAAGTTCAGTGGAAAG	qPCR
AS-BACH1-R2	TCCCTTTGTTGGAGTTTTGC	qPCR
AS-GAPDH-F1	TGTGGGCATCAATGGATTTGG	qPCR
AS-GAPDH-R1	ACACCATGTATTCCGGGTCAAT	qPCR
AS-RPL13A-F1	GCCCTACGACAAGAAAAAGCG	qPCR
AS-RPL13A-R1	TACTTCCAGCCAACCTCGTGA	qPCR
AS-ACTBL2-F1	GTCTGCCTTGGTAGTGGATAATG	qPCR
AS_ACTBL2-R1	TCGAGGACGCCCTATCATGG	qPCR
MA1033	GCCGCTAGAGGTGAAATTCTTG	qPCR
MA1034	CTTTCGCTCTGGTCCGTCTT	qPCR
TUBB F1	CTGGACCGCATCTCTGTGTA	qPCR
TUBB R1	TGCCCCAGACTGACCAAATAC	qPCR
ETS2 F1	CTCATGACTCCGCCAACTGT	qPCR
ETS2 R1	AGCCAGGGGTTCTTTGGAAT	qPCR
RUNX1 F1	CATCGCTTTCAAGGTGGTGG	qPCR
RUNX1 R1	ATGGCTGCGGTAGCATTTCT	qPCR
AIRE F1	CCCAGGCTCTCAACTGAAGG	qPCR
AIRE R1	GTCTGAATCCCGTTCCCGAG	qPCR
ERG F1	ATCGTGCCAGCAGATCCTAC	qPCR
ERG R1	AGAGAAGGATGTCGGCGTTG	qPCR
REST-v1 F1	GCACCCAACTTTACCACCCT	qPCR
REST-v1 R1	GGCCATAACTGTATTCGGCCT	qPCR
REST-v2 F1	CCGGCTGCGCGAATACAG	qPCR
REST-v2 R1	CAGGGCCATTCCAATGTTGC	qPCR
REST-v1/2 F1	GAGCTGGGGATAATGAGCGAG	qPCR
REST-v1/2 R1	GGGCGTTCTCCTGTATGAGTTC	qPCR

Primer Name	Sequence	Assay
UPS16 F	TACTGAGGCTCTTTCCGCAG	CHIP qPCR
USP16 R	TTGCAGCAGCTTAGACTCCA	CHIP qPCR
CD19 F	GCTGTTCCGGTGGAATGTTT	CHIP qPCR
CD19 R	CTTTGGCCCACACATACAGC	CHIP qPCR
NFKB1-1 F	AGAGGCACATGGGATTAGCG	CHIP qPCR
NFKB1-1 R	GTGAAGAGAAATGACCGCCG	CHIP qPCR
NFKB1-2 F	GATAAACCAGCAAGTCAGGGC	CHIP qPCR
NFKB1-2 R	TTATGCATGACCGCTGAAACA	CHIP qPCR
DYRK1A F	TGGTTACTTGCCTGTGCCTT	CHIP qPCR
DYRK1A R	TGGGCGGAATTCAAGAACCA	CHIP qPCR

Table 14-6 Human primer list

15 STRAINS USED IN THESE STUDIES

Strain Identifier	Well	Ploidy	Generation	WGS	aCGH	RNA Seq
PY5999	NA	1N	0	HiSeq2500	X	
PY6006	NA	2N	0	SOLiD	X	X
PY6040	NA	4N	0	SOLiD	X	X
1N_101	1N_A1	1N	250	HiSeq2500		
1N_102	1N_A6	1N	250	HiSeq2500		
1N_103	1N_A12	1N	250	HiSeq2500		
1N_104	1N_B2	1N	250	HiSeq2500	X	
1N_105	1N_C1	1N	250	HiSeq2500		
1N_106	1N_C5	1N	250	HiSeq2500		
1N_107	1N_C6	1N	250		X	
1N_108	1N_C9	1N	250	HiSeq2500		
1N_109	1N_C10	1N	250	HiSeq2500		
1N_110	1N_C12	1N	250	HiSeq2500	X	
1N_111	1N_D8	1N	250	HiSeq2500		
1N_112	1N_E1	1N	250	HiSeq2500		
1N_113	1N_E2	1N	250	HiSeq2500	X	
1N_114	1N_E3	1N	250	HiSeq2500	X	
1N_115	1N_E4	1N	250			
1N_116	1N_E11	1N	250	HiSeq2500		
1N_117	1N_F1	1N	250	HiSeq2500		
1N_118	1N_F5	1N	250	HiSeq2500		
1N_119	1N_F7	1N	250			
1N_120	1N_F8	1N	250	HiSeq2500		
1N_121	1N_F9	1N	250	HiSeq2500		
1N_122	1N_F11	1N	250	HiSeq2500		
1N_123	1N_G1	1N	250		X	
1N_124	1N_G2	1N	250	HiSeq2500		
1N_125	1N_G7	1N	250			
1N_126	1N_G12	1N	250			
1N_127	1N_H1	1N	250	HiSeq2500		
1N_128	1N_H6	1N	250	HiSeq2500		
1N_129	1N_H9	1N	250			
1N_131	1N_E8	1N	250	MiSeq	X	X
1N_132	1N_C11	1N	250	MiSeq	X	X
2N_201	2N_A10	2N	250		X	
2N_202	2N_B7	2N	250	HiSeq2500		
2N_203	2N_B11	2N	250	HiSeq2500		
2N_204	2N_B12	2N	250	HiSeq2500		

Strain Identifier	Well	Ploidy	Generation	WGS	aCGH	RNA Seq
2N_205	2N_C1	2N	250	HiSeq2500		
2N_206	2N_C2	2N	250	HiSeq2500		
2N_207	2N_C4	2N	250	HiSeq2500		
2N_208	2N_C8	2N	250	HiSeq2500		
2N_209	2N_D1	2N	250	HiSeq2500		
2N_210	2N_D4	2N	250	HiSeq2500		
2N_211	2N_D6	2N	250	HiSeq2500		
2N_212	2N_D7	2N	250	HiSeq2500	X	
2N_213	2N_E1	2N	250	HiSeq2500		
2N_214	2N_E2	2N	250	HiSeq2500		
2N_215	2N_E3	2N	250			
2N_216	2N_E4	2N	250			
2N_217	2N_E10	2N	250		X	
2N_218	2N_E12	2N	250	HiSeq2500	X	
2N_219	2N_F3	2N	250	HiSeq2500		
2N_220	2N_F4	2N	250			
2N_221	2N_F6	2N	250	HiSeq2500		
2N_222	2N_F9	2N	250	HiSeq2500		
2N_223	2N_F10	2N	250	HiSeq2500		
2N_224	2N_G4	2N	250	HiSeq2500		
2N_225	2N_G10	2N	250	HiSeq2500		
2N_226	2N_G11	2N	250	HiSeq2500		
2N_227	2N_H1	2N	250	HiSeq2500		
2N_228	2N_H2	2N	250		X	
2N_229	2N_H5	2N	250			
2N_230	2N_H7	2N	250		X	
2N_231	2N_H9	2N	250			
2N_232	2N_B2	2N	250	Solid	X	X
2N_233	2N_F12	2N	250	Solid	X	X
4N_301	4N_A1	4N	250		X	
4N_302	4N_A3	4N	250		X	
4N_303	4N_A7	4N	250			
4N_304	4N_B3	4N	250	HiSeq2500	X	
4N_305	4N_B4	4N	250	HiSeq2500	X	
4N_306	4N_B5	4N	250	HiSeq2500		
4N_307	4N_B6	4N	250	HiSeq2500		
4N_308	4N_B7	4N	250			
4N_309	4N_B9	4N	250	HiSeq2500		
4N_310	4N_B11	4N	250	HiSeq2500	X	
4N_311	4N_B12	4N	250	HiSeq2500		
4N_312	4N_C1	4N	250	HiSeq2500		
4N_313	4N_C3	4N	250	HiSeq2500	X	

Strain Identifier	Well	Ploidy	Generation	WGS	aCGH	RNA Seq
4N_314	4N_C11	4N	250	HiSeq2500		
4N_315	4N_D2	4N	250	HiSeq2500	X	
4N_316	4N_D3	4N	250	HiSeq2500	X	
4N_317	4N_D4	4N	250		X	
4N_318	4N_D8	4N	250		X	
4N_319	4N_E1	4N	250	HiSeq2500	X	
4N_320	4N_E2	4N	250	HiSeq2500	X	
4N_321	4N_E3	4N	250	HiSeq2500	X	
4N_322	4N_E8	4N	250	HiSeq2500		
4N_323	4N_E9	4N	250	HiSeq2500		
4N_324	4N_E10	4N	250	HiSeq2500		
4N_325	4N_F1	4N	250			
4N_326	4N_F4	4N	250			
4N_327	4N_F6	4N	250	HiSeq2500		
4N_328	4N_F10	4N	250	HiSeq2500		
4N_329	4N_F11	4N	250	HiSeq2500	X	
4N_330	4N_G4	4N	250		X	
4N_331	4N_H6	4N	250	HiSeq2500		
4N_332	4N_H10	4N	250	HiSeq2500		
4N_333	4N_H11	4N	250	HiSeq2500		
4N_334	4N_G11	4N	250	SOLiD	X	X
4N_335	4N_G2	4N	250	SOLiD	X	X
4N_336	4N_A8	4N	250	SOLiD	X	X
4N_337	4N_F2	4N	250	SOLiD	X	X
4N_g500_401	4N_g500_A1	4N	500		X	
4N_g500_402	4N_g500_A3	4N	500		X	
4N_g500_406	4N_g500_B5	4N	500		X	
4N_g500_407	4N_g500_B6	4N	500		X	
4N_g500_418	4N_g500_E3	4N	500		X	
4N_g500_420	4N_g500_E8	4N	500		X	
4N_g500_428	4N_g500_F11	4N	500		X	
4N_g500_431	4N_g500_G4	4N	500		X	
4N_gen35_501	4N_gen35_A3	4N	35		X	
4N_gen55_502	4N_gen55_A3	4N	55		X	
4N_gen35_503	4N_gen35_G2	4N	35		X	
4N_gen55_504	4N_gen55_G2	4N	55		X	
4N_gen55_505	4N_gen55_A8	4N	55		X	
4N_gen55_506	4N_gen55_A12	4N	55		X	
4N_gen55_507	4N_gen55_F11	4N	55		X	
4N_gen55_508	4N_gen55_H5	4N	55		X	

Table 15-1 Evolved strains used in these studies

Strain	Parental Strain	Relevant genotype	Source
BY4742	S288c	*MATα his3Δ lys2Δ leu2Δ ura2Δ*	Dowell Collections
BY3295	BY4741	*MATa his3Δ leu2Δ met15Δ ura2Δ*	(Selmecki et al. 2015)
PY5997	BY3295	*matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR*	(Selmecki et al. 2015)
PY5998	PY5997	**1N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-ceCFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY5999	PY5997	**1N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-eYFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY6006	PY5999	**2N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-eYFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY6008	PY5998	**2N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-ceCFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY6014	PY5999	**2N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-eYFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY6022	PY5998	**2N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-ceCFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY6031	PY6008	**4N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-ceCFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY6032	PY6022	**4N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-ceCFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY6040	PY6006	**4N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-eYFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY6045	PY6014	**4N** *matΔ::pSTE5-ura3::TRP1, ste4Δ::HygroR, trp1::NatR::pGAL-eYFP-tADH-SpHIS5*	(Selmecki et al. 2015)
PY7284	PY5999	**1N** *SNF3-G439E*	(Selmecki et al. 2015)
PY7289	PY5999	**2N** *SNF3-G439E/SNF3*	(Selmecki et al. 2015)
PY7290	PY5999	**2N** *SNF3-G439E/SNF3*	(Selmecki et al. 2015)
PY7285	PY5999	**2N** *SNF3-G439E/SNF3-G439E*	(Selmecki et al. 2015)
PY7286	PY5999	**2N** *SNF3-G439E/SNF3-G439E*	(Selmecki et al. 2015)
PY7287	PY5999	**2N** *SNF3-G439E/SNF3-G439E*	(Selmecki et al. 2015)
PY7288	PY5999	**2N** *SNF3-G439E/SNF3-G439E*	(Selmecki et al. 2015)
PY7291	PY5999	**4N** *SNF3-G439E/SNF3/SNF3/SNF3*	(Selmecki et al. 2015)
PY7292	PY5999	**4N** *SNF3-G439E/SNF3/SNF3/SNF3*	(Selmecki et al. 2015)
PY7293	PY5999	**4N** *SNF3-G439E/SNF3/SNF3/SNF3*	(Selmecki et al. 2015)
PY7294	PY5999	**4N** *SNF3-G439E/SNF3/SNF3/SNF3*	(Selmecki et al. 2015)
PY7295	S288c	**2N** RLY4737 *MATa/α ura3Δ his3Δ trp1Δ leu2Δ*	(Pavelka, Rancati, Zhu, et al. 2010)
PY7296	PY7295	**2N** RLY4888 *MATa/α* + chr. XIII trisomy	(Pavelka, Rancati, Zhu, et al. 2010)
PY7297	PY7295	**4N** *MATa/a/α/α*	(Selmecki et al. 2015)
PY7298	PY7295	**4N** *MATa/a/α/α*	(Selmecki et al. 2015)
PY7299	PY7295	**4N** *MATa/a/α/α*	(Selmecki et al. 2015)
PY7300	PY7295	**4N** *MATa/a/α/α* + chr. XIII pentasomy	(Selmecki et al. 2015)
PY7301	PY7295	**4N** *MATa/a/α/α* + chr. XIII pentasomy	(Selmecki et al. 2015)
PY7302	PY7295	**4N** *MATa/a/α/α* + chr. XIII pentasomy	(Selmecki et al. 2015)
PY7303	PY7295	**4N** *MATa/a/α/α* + chr. XII pentasomy	(Selmecki et

Strain	Parental Strain	Relevant genotype	Source
			al. 2015)
PY7304	PY7295	**4N** *MATa/a/α/α* + chr. XII pentasomy	(Selmecki et al. 2015)
PY7305	PY7295	**4N** *MATa/a/α/α* + chr. XII pentasomy	(Selmecki et al. 2015)
ASY001	PY5999	**1N** *rgt1Δ::KanMX*	This study
ASY002	PY5999	**1N** *rgt1Δ::KanMX*	This study
ASY003	PY5999	**1N** *rgt1Δ::KanMX*	This study

Table 15-2 Engineered strains used in these studies