Molecular Mechanisms Controlling Complex Traits in Yeast

by

Brian L. Chin

B.A. Molecular and Cellular Biology
Univeristy of California, Berkeley, 2002

Submitted to the Department of Biology in partial fulfillment
of the requirements for the thesis degree of

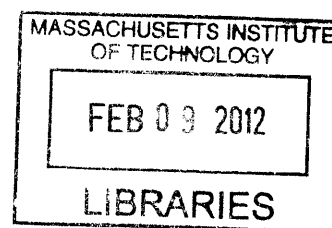Doctor of Philosophy in Biology

at the

Massachusetts Institute of Technology

February 2012
[June 2012]

Signature of Author: _____

Department of Biology

Certified by: _____

Gerald R. Fink, Ph.D.
Herman and Margaret Sokol Professor, Department of Biology
Thesis Supervisor

Accepted by: _____

Alan D. Grossman, Ph.D.
Praecis Professor of Biology
Chairman, Committee for Graduate Students

Molecular Mechanisms Controlling Complex Traits in Yeast

by

Brian L. Chin

Abstract

A fundamental goal in biology is to understand how the information stored in DNA results in a cellular function. However, it is insufficient to study one variant of a particular DNA sequence because most people do not share identical genome sequences, and the differences in DNA sequence have functional consequences. In this thesis, I examine how natural variation in the *Saccharomyces cerevisiae* genome can affect cellular processes. This is done using deletion libraries to examine how mutations in the same gene but in two different genetic backgrounds of *S. cerevisiae*, S288c and Σ1278b, can lead different phenotypes for two traits: gene essentiality and agar adhesion. We found that the genomes of the S288c and Σ1278b strains are only as divergent as two humans in the population. However, analyses of deletion libraries in each strain revealed 57 genes have functions that are essential in one strain but not the other. Strain specific phenotypes are more pronounced for the trait of agar adhesion where 553 deletions have phenotypes that are specific to one strain or the other. Part of the difference is because the Σ1278b strain requires the filamentation mitogen activated kinase pathway (fMAPK) for agar adhesion but the S288c strain does not. I found that S288c is able to bypass the fMAPK pathway because it contains an allele of the transcription factor *RPI1* that promotes transcription of the gene *FLO11*. Characterization of the sequence differences between the S288c and Σ1278b alleles of *RPI1* revealed that they differ in the number of intragenic tandem repeats. Examination of the genomes of both strains uncovered the possibility that expansions and contractions of intragenic repeats may be a general mechanism to quickly introduce genomic and phenotypic variation.

Thesis Advisor: Gerald R. Fink, PhD
Title: Herman and Margaret Sokol Professor, Department of Biology

Acknowledgements

This dissertation was only possible because of the support and guidance that I have received during my graduate career. First I need to thank my thesis advisor Dr. Gerald R. Fink for his professional guidance and for providing an exceptional research environment. It has been a privilege to learn from his vast experience and I hope that I have gained a fraction of his intuition for biological processes.

Many members of the Fink lab helped me throughout my graduate career. Kevin Verstrepen, Validmir Voynov and Stacie Bumgarner all provided valuable assistance in learning yeast biology and techniques. Paula Grisafi and Anna Symbor-Nagrabska have been exceptional bay mates and allowed me to peak into the intricacies of managing a lab. William Timberlake and Margaret Frizzell have been outstanding collaborators and helped me to realize the FASTER MT method.

My committee members Frank Solomon and Chris Kaiser have been invaluable during my time at MIT. They have been generous in providing both scientific advice and career advice. Their insights were instrumental to the progress of my work.

I also need to thank various collaborators outside of the Fink lab that I have had the opportunity to work with. David Gifford, Robin Dowell, and Fran Lewitter helped me to examine the genome sequence of Sigma. Also, Charlie Boone and Owen Ryan built the Sigma deletion library and were very generous with reagents and data pertaining to the library.

My research career and love for genetics started during my time as an undergraduate at the University of California at Berkeley. I am thankful to Dr. Krishna K. Niyogi, a former member of the Fink lab, for taking me in as an undergraduate researcher and providing me with my own project to solve. In his lab I was able to work with Rachel Dent and Heidi Ledford who helped me to obtain my first taste for genetic analyses, designing screens and selections, and manipulating genomic data.

Finally I would like to thank my family and friends who have supported me throughout my time at MIT.

Table of Contents

Chapter 1

Introduction

Functional variation due to sequence differences is central to our understanding of several aspects of biology. In evolution, how organisms evolve to deal with new environmental challenges relies upon modifying current functions to take on new roles. These modifications occur constantly and, in the human population, these alterations in DNA sequence are responsible for many genetic diseases. These alterations in DNA sequence within a species have been studied on a population basis, but they have not been subjected to a genome-wide analysis. To address the extent and consequences of functional differences in a genome, this thesis characterizes the natural variation between two closely related strains of yeast and the functional consequences for biological processes in each strain.

In the introduction to this thesis I begin by briefly describing studies that have examined genomic variation and its impact on phenotypic variation. I first mention the difficulties faced in studying natural variation in the human population. Subsequently, I present previous reports of natural variation performed in the budding yeast *S. cerevisiae* with an emphasis on the prospects for using yeast to understand how natural variation impacts a complex trait. Finally, I conclude this chapter by introducing the complex regulation governing the expression of the *FLO* (*FLO*cculin) gene family, which is the basis for the work presented in the later chapters.

In Chapter Two, I present an initial genome-wide study comparing two strains of *S. cerevisiae*, S288c and Σ1278b (Sigma). This study includes the high quality sequence of the Sigma genome and its comparison to the S288c genome. I then describe the construction of a

7

Sigma deletion collection in which every non-essential ORF was systematically deleted. This deletion library was constructed by methods similar to those previously used to create a deletion library in the S288c strain background and the existence of the two libraries permits the comparison of phenotypes in cells of different genetic backgrounds.

In Chapter Three, I present the results of screens performed on the two deletion libraries to examine yeast adhesion properties within a strain and between strains. The screens identified many strain specific regulators of adhesion. The most salient of these results is the finding that a mitogen activated kinase (MAPK) pathway has strain specific effects on adhesion and *FLO11* gene expression. Using a selection, the S288c allele of the transcription factor *RPI1* was discovered to permit transcription of the *FLO11* gene in the absence of the function of the filamentation MAPK (fMAPK) pathway. The characterization of this allele revealed that it is an allele-specific transcriptional activator that participates in *FLO11* expression.

In Chapter Four, I summarize these finding and discuss how the methods and findings reported here will impact future studies of natural variation.

Natural Variation in the human population

The human population is very diverse with recognizable differences in traits such as height, girth, and pigmentation. While these outward manifestations of phenotypic diversity are the easiest to measure, there is also diversity in traits that develop with time and appear more complex . Some of these complex traits include the predisposition to different diseases, such as type 2 diabetes, heart disease, and cancer. The genetic makeup of different individuals plays an

important role in the diversity of these traits in the population and the understanding of this genetic diversity will be necessary for the treatment and prevention of many genetic diseases.

A variety of methods have been utilized by human geneticists to determine the casual polymorphisms for a number of different genetic diseases. Familial linkage studies were successful in locating the casual variants for monogenic diseases such as sickle cell anemia, cystic fibrosis, and Huntington's disease (GROUP 1993; KEREM *et al.* 1989; RIORDAN *et al.* 1989; ROMMENS *et al.* 1989). However, pedigree analysis proved to have insufficient statistical power to elucidate the inheritance of many multi-genic diseases.

It was hoped that these multi-genic traits could be understood with the help of the human genome sequence. The human genome sequence allowed for the discovery of many polymorphic markers. With modern microarray and sequencing technology it was possible to detect the presence of these markers and their association with disease across large numbers of individuals. If a DNA polymorphism is identified as common among a group of individuals manifesting the disease, then that marker is linked to the same piece of DNA as a disease causing mutation (Figure 1-1). Studies using this methodology have been termed Genome Wide Association Studies (GWAS) (ALTSHULER *et al.* 2008). Unfortunately, the ensemble of polymorphisms responsible for these multi-genic disease states is complex and makes associations using GWAS difficult.

One of these limitations in using associated polymorphisms to uncover loci in the genome is the lack of resolution. The use of associated polymorphisms allows the general region of the chromosome associated with a trait to be identified but the causal polymorphism can be one of many in a large chromosomal region. While GWAS studies often result in regions of only

10-100 kb in size, many of these regions have not been thoroughly resolved to locate the causal mutation (IOANNIDIS *et al.* 2009). In the absence of exact determination of the responsible gene, it is not possible to know the mechanism behind how the region impacts the trait.

In addition, even where the statistical resolution has been sufficient, the loci discovered through GWAS only explain a minority of the heritability of the trait (MANOLIO *et al.* 2009). This failure to explain the heritability has been termed "the missing heritability." Studies of the genetic inheritance of height provide an example of this missing heritability. One GWAS identified 20 loci associated with this trait yet this large number of loci is predicted to explain only 3% of the heritability of height variation (WEEDON *et al.* 2008).

Moreover, particular DNA variants that have been associated with a disease have poor predictive power despite having significant levels of association (JAKOBSDOTTIR *et al.* 2009). Therefore, not only are researchers are unable to find all the casual variants but the variants that have been found do not predict the risk for the disease. This lack of predictability is a major obstacle to the implementation of "personalized medicine" – the hope that treatments can be tailored to the individual patient's genotype. Personalized medicine depends upon being able to determine an individual's phenotype based on their genotype, and has been successful in single Mendelian traits where few genes are involved. However, it is still the research hope for genetically complex diseases.

Given the large amount of effort being placed into GWAS, why do these problems exist? First, the inability to locate most of the casual variants of a trait may be due to technical limitations where too few individuals or too few SNPs have been examined to find all of the variants. Additionally, studies often separate the predicted causes of a disorder into

environmental and genetic components. This way of looking at heritability in GWAS may be fundamentally flawed (VINEIS and PEARCE 2011). It is possible to envision disorders where the contribution of the environment is inseparable from the genetic component. For example, certain alleles may predispose an individual to heart disease only under a certain diet. For this reason, Vineis and Pearce (2011) propose that heritability cannot be separated into genetic and environmental effects unless all individuals are experiencing the same environment.

Second, there may be complex interactions between the many polymorphisms that contribute to disease. One can imagine "And" interactions in which several different polymorphisms must be present to yield the final phenotype and "Or" polymorphisms in which any one of a number of different polymorphisms is sufficient to contribute to the phenotype. Thus, the inability to link disease variants to an individual's disease risk may be due to an insufficient understanding of the molecular mechanisms underlying the variants and how the variants interact with each other. The interaction between variants or "epistasis" (where the phenotypic effect of one locus depends on the genotype at one or more different loci) has been suggested by a number of authors as an important issue (CARLBORG and HALEY 2004; KROYMANN and MITCHELL-OLDS 2005). If gene interaction is important in the outcome of a phenotype, then a single DNA variation may contribute to a phenotype in different ways depending on the other collection of polymorphisms in the individual.

It is difficult to address these issues when examining the human population where controlled crosses cannot be used to identify the gene interactions. Sample sizes can be increased to try and gain more statistical power for finding relevant loci, and ultimately some association of various combinations of polymorphisms may be predictive. Since a uniform

environment cannot be provided for humans, such studies will inevitably have variation between identical genotypes.

In contrast to human studies, these problems can be overcome when working in model organisms in the laboratory. Examining natural variation in model organisms will have the added benefit of providing an understanding of the basic biology behind gene interactions as well as insights into how to approach natural variation in the human population.

Natural variation in *S. cerevisiae*

The use of *S. cerevisiae* to examine natural variation has many advantages. Cells can be grown in a uniform environment, controlled crosses are easy to perform, and methods exist to obtain large quantities of meiotic progeny for mapping studies. Determination of variants is also facilitated by yeast's compact genome and efficient gene knockout and replacement techniques. Because of these advantages, linkage studies in *S. cerevisiae* have been successful in finding some determinants for a variety of different traits and some of the more salient of these studies will be described in the ensuing paragraphs.

The advantages of using *S. cerevisiae* to identify the effects of natural variation are illustrated by a cross of two *S. cerevisiae* strains from the Kruglyak lab (BREM *et al.* 2002). They crossed the common laboratory strain S288c (BY) by a wild strain isolated from a wine barrel, RM11 (RM). Markers in over 100 progeny from the BY x RM cross have been mapped with high resolution using microarrays. These progeny were subsequently examined for loci linked to differences in a variety of traits including gene expression, small molecule sensitivities, cell-cell aggregation, telomere length, morphological variation and sensitivity to DNA-damaging

agents (BREM and KRUGLYAK 2005; DEMOGINES *et al.* 2008; GATBONTON *et al.* 2006; NOGAMI

et al. 2007; PERLSTEIN *et al.* 2006; RONALD *et al.* 2005; YVERT *et al.* 2003). These studies

provide an important look into the challenges and benefits of using mapping methods to examine

functional variation.

The study of gene expression variation in the BY x RM cross gave insights into the

molecular mechanisms governing some gene expression differences between the two strains

(RONALD *et al.* 2005). In this cross, 12-20% of genes with expression differences have variation

in cis-regulatory elements that cause variation in transcript abundance (RONALD *et al.* 2005).

Several trans-acting factors were also discovered, including a variant of the *AMN1* gene that

enhances mother-daughter cell separation after cytokinesis and a variant of *GPA1* that enhances

mating (YVERT *et al.* 2003).

The study of resistance to DNA damaging agents among the BY x RM progeny provided

insight into the utility of controlled crosses and gene replacement in *S. cerevisiae* in dissecting a

complex trait (DEMOGINES *et al.* 2008). Although both the BY and RM parent strains are

resistant to the DNA damaging agent 4-nitroquinoline 1-oxide (4-NQO), some progeny from this

cross were sensitive to 4-NQO. Using linkage mapping, Demogines et al (2008) found that the

RM allele of *RAD5* contributed to the sensitivity, but since the RM parent was resistant, other

alleles from the BY strain must interact with the RM *RAD5* to lead to a 4-NQO sensitive

phenotype. Using molecular techniques, the researchers then replaced the BY *RAD5* with the

RM *RAD5*. This modified BY strain was then used in backcrosses to a standard RM strain.

While the segregation pattern of 4-NQO sensitivity remained complex through the third

backcross, *MKT1* was uncovered as another gene affecting 4-NQO through mapping methods on

13

this generation. The fact that *MKT1* could only be uncovered from the backcrossed generation and not from their original linkage study suggests that many more loci are present in the original cross that can modify the polymorphism of *MKT1* associated with 4-NQO sensitivity. It was only by eliminating those other loci through controlled crosses could the researchers obtain other modifiers.

The BY x RM cross also provided the basis for X-QTL, an enhanced mapping method in yeast. X-QTL vastly increases the statistical power of linkage studies in *S. cerevisiae* while simultaneously decreasing the cost and effort associated with genotyping the progeny (EHRENREICH *et al.* 2010).

X-QTL has the potential to increase the power of mapping studies while simultaneously streamlining the process because it couples array assisted bulk segregant analysis with genetic markers to select large pools of meiotic progeny. In array assisted bulk segregant analysis, instead of genotyping individual progeny, pools of progeny of extreme phenotypes (e.g. resistant to a drug) are genotyped for relative marker levels using microarrays (BRAUER *et al.* 2006). A polymorphism that is either the causal variant or is linked to the causal variant will be enriched in the pool whereas unlinked markers will be present at equal frequencies. Thus, finding QTLs is reduced from genotyping hundreds of progeny to genotyping a single pool of progeny. Most initial bulk segregant analyses used pools of a few hundred progeny and successfully mapped loci associated with in variety of traits, including auxotrophies, growth defects on acetate, flocculation, adaptation to fluctuating carbon sources, resistance to DNA-damaging agents, and resistance to leucine starvation (BOER *et al.* 2008; BRAUER *et al.* 2006; DEMOGINES *et al.* 2008; SEGRE *et al.* 2006).

Although examination of a few hundred progeny is sufficient to find associations in traits with simple inheritance patterns, the mapping of more complex, multi-genic traits requires thousands of progeny to find associations for loci with modest effects (EHRENREICH *et al.* 2010). X-QTL leverages markers developed to select for meiotic progeny in yeast to obtain pools consisting of thousands of meiotic progeny (EHRENREICH *et al.* 2010; TONG and BOONE 2006). This entire pool is then subjected to a second selection for the trait of interest (e.g. resistance to DNA-damaging agents) and genotyped. This use of extremely large pools allows for the detection of over a dozen loci with significant association to a trait (EHRENREICH *et al.* 2010). X-QTL has been successfully used to uncover loci affecting differences among BY x RM progeny for sensitivity to small molecules (e.g. ethanol, SDS, and NaCl), mitochondrial function, and translation termination (EHRENREICH *et al.* 2010; TORABI and KRUGLYAK 2011).

Despite the vast amount of data generated from the studies of the BY x RM cross, all of the studies emphasize that there is more complexity in the genetic control of these traits than is fully explained by the polymorphisms uncovered. For example, Nogami et al (2007) examined 501 different morphological traits and found 143 traits differ between the BY and RM parents, but they could only uncover statistically significant QTLs for 27 of the traits. Notably, 104 traits did not differ between the parents but did differ among the F1 progeny (Figure 1-2). One explanation for this phenomenon is that the parents possess alleles in different genes that act in opposing fashions, and when they recombine in the cross it results in transgressive segregation, where the progeny show more extreme phenotypes than the parents. Transgressive segregation requires differences in multiple loci controlling that trait, but despite finding evidence for transgressive segregation in 34 traits, only 12 traits had confirmed linkage to multiple identified

loci (NOGAMI *et al.* 2007). This result suggests that many other loci remain to be detected and many of the interactions are unidentified.

These studies have resulted in the mapping of many different QTLs in *S. cerevisiae*, but the molecular mechanisms linking the variant to the trait are often left unexplored. The absence of a functional explanation is akin to the analysis of human pedigrees where loci affecting a trait can be uncovered, but the mechanisms of gene interactions affecting the trait remain unknown.

Studies of natural variation in sporulation efficiency of two different yeast strains illustrates the difficulties faced in finding molecular mechanisms from mapping data and the potential myriad of phenotypes among the progeny of a cross. Four nucleotide changes in the three genes *RME1*, *IME1*, and *RSF1* were found to explain the majority of the difference in sporulation efficiency between the inefficiently sporulating vineyard strain BC187 (V) and the efficiently sporulating oak strain YPS606 (O) (Gerke et al. 2009). A thorough epistasis analysis of the possible allele combinations in the progeny of a cross of V x O revealed that the alleles act synergistically to affect sporulation efficiency. Based upon the known functions of the genes, a model was formed that explains the complexity of the epistasis data (Figure 1-3) (GERTZ *et al.* 2010). The model centers on Ime1p, which is a transcription factor that promotes sporulation. Rme1p represses *IME1* transcription while Rsf1p promotes it.

The phenotypes of the progeny, which vary enormously in their ability to promote sporulation, are best explained by the following. The V strain has a cis-acting promoter allele of *RME1* that leads to increased expression of Rme1p, which causes repression of sporulation. The O strain has two polymorphisms in the *IME1* gene: a promoter allele that decreases the binding of Rme1p and a change in the *IME1* ORF that leads to greater activity. The net effect of these

two mutations is to enhance Ime1p production and therefore sporulation. In addition, the O

strain has a polymorphism in the *RSF1* ORF that enhances its stimulation of *IME1* transcription.

These four polymorphisms segregate in a cross between O and V to produce progeny with a

spectrum of sporulation capabilities that required considerable backcrossing and analysis with

high resolution. Even with this detailed analysis it would be difficult to predict the exact

sporulation capability of progeny given the genotypes at these four loci.

This study of epistasis between sporulation QTLs provides a unique look into the

molecular mechanisms behind a complex trait. This complex level of gene-gene and gene-DNA

interactions has not been worked out for any other traits, yet studies have already suggested that

this type of gene interaction makes the association of genotype with phenotype in humans

difficult as one cannot breed the combinations of alleles that are informative for a functional

assessment (CARLBORG and HALEY 2004; KROYMANN and MITCHELL-OLDS 2005).

The molecular mechanisms underlying the natural variation in yeast sporulation were

only able to be uncovered because a significant amount was known about regulatory control of

meiosis. Future studies of natural variation in yeast will also benefit if they examine traits with

well-characterized regulation. In addition to having all the interacting genes identified, an ideal

system for analysis should also have the complex regulatory circuits understood. One potential

network that meets both criteria is the system of genes controlling the expression of the yeast cell

surface proteins. Many factors affecting the structure of the cell surface have been determined,

and significant phenotypic variability is known to exist (KLIS 1994; SMITS *et al.* 1999;

VERSTREPEN *et al.* 2005).

Regulation of *FLO* gene expression

The outer surface of microbes plays a critical role in communicating and interacting with the environment. Changes in the cell surface can promote adhesion to substrates, biofilm formation and evasion of a host immune system (REYNOLDS and FINK 2001). Therefore, the proper control over the components of the cell surface is necessary for the survival and adaptation of a microbe to ever changing environments. The yeast *S. cerevisiae* is capable of changing a number of properties of its cell surface, but adhesiveness is one of the most readily observable changes.

Adhesion in *S. cerevisiae* is mediated by the *FLO* gene family of cell surface adhesion proteins. In the reference strain, S288c, there are five members of this family, *FLO1*, *FLO5*, *FLO9*, *FLO10*, and *FLO11*. These genes all share a similar structure with three distinct domains: a C-terminal domain which anchors the protein to the cell surface, heavily glycosylated central domain that is often rich in serine and threonine repeats; and an N-terminal domain that sometimes shows lectin-like binding to certain sugars or peptides (LO and DRANGINIS 1996; STRATFORD and ASSINDER 1991).

Despite the similar structures of the FLO genes, expression of different members of the *FLO* gene family confer different properties to the cell (GUO *et al.* 2000). Flo11p expression in diploid cells is necessary for a developmental switch from a yeast form cell to a filamenteous form (LO and DRANGINIS 1998). In the absence of Flo11p expression, diploid yeast grow as ovoid cells, but upon Flo11p expression the cells can switch to a filamenteous form where the cells increase in length and change the polarity of their budding (KRON *et al.* 1994). This switch

occurs under situations of nutrient deprivation and is thought to permit the usually sessile cells to spread into new environments (GIMENO *et al.* 1992).

Like diploid cells, haploid cells undergo a similar switch (haploid invasive growth), although the morphological changes differ slightly (ROBERTS and FINK 1994). In particular, the elongation is not as pronounced and Flo11p expression in haploid cells allows the yeast to adhere well to different substrates, including plastics and agar.

In contrast to Flo11p, expression of Flo1p does not facilitate filamentation or cell-substrate adhesion but it leads to cell-cell adhesion (GUO *et al.* 2000). When a culture of cells expressing Flo1p is grown in liquid media, either in the lab or in an industrial fermentation, the cells will aggregate into clumps. This cell-cell adhesion process is called flocculation and has been heavily utilized in the brewing industry to remove the yeast after fermentation because the clumps of cells readily sediment. In addition, Flo1p mediated aggregation has been shown to confer resistance to different chemical stresses, including ethanol and the anti-fungal agent amphotericin B (SMUKALLA *et al.* 2008).

The genetic regulation of FLO gene expression is best understood for *FLO11* where a complex network of at least four parallel signaling pathways participate in activating transcription of *FLO11*: PKA, SNF, TOR and MAPK (Figure 1-4) (BRUCKNER and MOSCH 2012; CULLEN and SPRAGUE 2012; VERSTREPEN and KLIS 2006). Central to the PKA pathway are the three protein kinase A proteins in *S. cerevisiae*: Tpk1p, Tpk2p, and Tpk3p. In presence of cAMP, the regulatory protein Bcy1p dissociates from the Tpk proteins to activate their kinase activity (PAN and HEITMAN 1999). Despite a shared mode of regulation, the different Tpk proteins have different roles in *FLO11* regulation. Tpk3p appears to have an inhibitory effect on

FLO11 expression, but the mechanism of inhibition is not yet known. Tpk2p promotes *FLO11* expression by phosphorylating and activating the transcription factor Flo8p while simultaneously phosphorylating and inactivating the repressor Sfl1p (Robertson and Fink 1998, Pan and Heitman 2002, Furukawa et al 2009). Tpk1p appears to repress *FLO11* expression by inhibiting the kinase Yak1p, which in turn inhibits the transcriptional repressor Sok2p (MALCHER *et al.* 2011). Sok2p in turn represses both *FLO11* and the transcriptional activator Phd1p (BORNEMAN *et al.* 2006; GIMENO and FINK 1994). The receptor for the PKA pathway is the G-protein coupled receptor Gpr1p. Gpr1p is a sugar sensing receptor that works with the Gα subunit Gpa2p to activate the PKA pathway (Lorenz et al 2000, Harashim et al 2006, Thevelein and Voordeckers 2009).

In addition to the PKA pathway, *FLO11* expression requires the filamentation MAPK (fMAPK) pathway. Genetic studies have suggested that the receptor for the fMAPK pathway is the cell surface protein Msb2p; however, the ligand for Msb2p remains unknown (PITONIAK *et al.* 2009). Msb2p is necessary for the eventual activation of a canonical MAPK signaling cascade consisting of the MAPKKK Ste11p, the MAPKK Ste7p and the MAPK Kss1p. Kss1p acts in an unphosphorylated state as an inhibitor of the *FLO11* expression by binding to the transcription factor Ste12p in coordination with the negative regulators Dig1p and Dig2p (COOK *et al.* 1996; COOK *et al.* 1997). The binding of Kss1p, Dig1p and Dig2p to Ste12p inhibits Ste12p's ability to interact with another transcription factor, Tec1p, and the Ste12p-Tec1p interaction is necessary for this combined oligomer to bind to the *FLO11* promoter and activate *FLO11* transcription. Upon activation of the fMAPK pathway, Kss1p phosphorylates Dig1p and Dig2p, relieving repression of Ste12p which can then interact with Tec1p to activate *FLO11*

expression (CHOU *et al.* 2006; KOHLER *et al.* 2002; MADHANI and FINK 1997; ZEITLINGER *et al.* 2003).

The fMAPk pathway shares many of the kinases with the pheromone response pathway, namely Ste20p, Ste11p, Ste7p, and Ste12p (LIU *et al.* 1993). In addition, Ste20p and Ste11p also participate in the high osmolarity glycerol (HOG) pathway. Despite these shared components, the fMAPK, pheromone response and HOG pathways are all able to activate distinct sets of genes (O'ROURKE and HERSKOWITZ 1998).

The TOR pathway and the Snf1 pathway are two additional pathways that regulate *FLO11*. The TOR pathway is important in sensing nitrogen and acts through the transcription factor Gcn4p to regulate *FLO11* (BRAUS *et al.* 2003; VINOD *et al.* 2008). The Snf1 pathway controls *FLO11* in response to glucose by regulating the repressors Nrg1p and Nrg2p at the *FLO11* promoter (KUCHIN *et al.* 2002; VYAS *et al.* 2003).

In addition to the previously mentioned signal transduction pathways, *FLO11* is also subject to epigenetic regulation where a genetically homogenous population of yeast contains a mixture of cells where some express Flo11p and others will not. This variegation in expression and heterogeneity in Flo11 transcription occurs in spite of the cells being exposed to identical environments (BUMGARNER *et al.* 2009; HALME *et al.* 2004; OCTAVIO *et al.* 2009). The variegated expression of Flo11p is due to regulation by several chromatin modifying factors, such as the histone deacetylase *HDA1*, that contribute to determining whether the promoter of *FLO11* is in a transcriptionally repressed state or a transcriptionally competent state.

In addition to protein factors that determine the chromatin state at the *FLO11* promoter, Flo11p's variegated expression is also controlled by cis-interfering noncoding RNAs. These

noncoding RNAs toggle the *FLO11* promoter between a transcriptionally repressed state and a transcriptionally competent state (BUMGARNER *et al.* 2009). The presence of these RNAs helped to explain the finding that the Rpd3L histone deacetylase complex is needed to maintain the *FLO11* promoter in a competent state. Histone deacetylases are normally associated with compaction of DNA into heterochromatin and this compaction is inhibitory toward transcription (BERGER 2007). Therefore, the finding that Rpd3L acted directly on the *FLO11* promoter and but was needed for the transcription competent state seemed contradictory to its function. This paradox was solved with the finding that Rpd3L repressed the expression of the noncoding RNA *ICR1* in the *FLO11* promoter. Transcription of *ICR1* leads to an inactive *FLO11* promoter, but the Rpd3L complex can repress *ICR1* transcription and allow for a switch to a transcriptionally competent *FLO11* promoter (BUMGARNER *et al.* 2009).

While studies of *FLO11* expression have uncovered many regulatory mechanisms, the regulatory control over many other *FLO* genes has not received the same level of attention. *FLO1* is the next best studied *FLO* gene and like *FLO11*, its regulation is dependent upon the transcription factor *FLO8* (KOBAYASHI *et al.* 1999). *FLO1* is also regulated by different chromatin modifiers. As with *FLO11*, *FLO1* expression can be repressed by the histone deacetylase Hda1p, although it is not known if Hda1p acts directly at the *FLO1* promoter (DIETVORST and BRANDT 2008). However, unlike Flo11p expression, Flo1p expression can also be regulated by the COmplex Proteins Associated with Set1 (COMPASS) methylation complex. The COMPASS complex methylates histone H3 (KROGAN *et al.* 2002). This activity is associated with both general transcription activity and silencing at specific loci (BOA *et al.* 2003; KROGAN *et al.* 2002).

These reports detailing *FLO1* regulation have come from analysis of regulation in a variety of genetically different *S. cerevisiae* strain backgrounds. Some of the strains were modified laboratory strains (KOBAYASHI *et al.* 1999), and others are of unknown, but possibly industrial, backgrounds (DIETVORST and BRANDT 2008; FLEMING and PENNINGS 2001). The use of these strains is necessary because, with one exception, most laboratory strains do not express any of the *FLO* genes. The exception is the Sigma strain which is competent for expression a single FLO gene, *FLO11* (GUO *et al.* 2000). This property has made Sigma the default strain for the study of filamentous growth, agar adhesion and *FLO11* expression

The S288c strain is one of the lab strains that does not normally expresses any of the *FLO* genes, but it is a better characterized strain than Sigma. S288c does not flocculate or adhere to agar because of a nonsense mutation in the *FLO8* transcription factor that prevents expression of *FLO11* and *FLO1* (LIU *et al.* 1996). In an S288c *FLO8* strain *FLO11* is expressed and it will adhere to agar but, unlike Sigma, S288c *FLO8* will also express *FLO1* and flocculate (KOBAYASHI *et al.* 1996; KOBAYASHI *et al.* 1999; LIU *et al.* 1996).

S288c and Sigma: do little differences matter?

Many lines of evidence have suggested that S288c and Sigma are closely related strains. Both strains share a common set of progenitors (MORTIMER and JOHNSTON 1986), and they can mate and produce viable progeny at high frequency. Previous microarray based genotyping studies have suggested that S288c and Sigma are more closely related than either is to RM11, SK1, or YJM789, which are other strains that are frequently used in natural variation studies (SCHACHERER *et al.* 2009; WINZELER *et al.* 2003).

Despite the proposed genomic similarities, a few differences between the two strains have been observed. Originally, Sigma was studied because it represses expression of the *GAP1* amino acid permease in the presence of ammonia whereas the S288c strain does not (RYTKA 1975). This difference made Sigma the strain of choice for studies of *GAP1* regulation, and later studies have found other differences between Sigma and S288c.

Sigma has two functional aquaporins, *AQY1* and *AQY2*, which are more common in wild and industrial strains than in laboratory strains (LAIZE *et al.* 2000). S288c has nonfunctional alleles of these genes and the presence of these alleles in Sigma improves Sigma's freeze tolerance (TANGHE *et al.* 2002). Additionally, many laboratory strains have a tandem array of plasma membrane Na+-ATPase exporters. S288c has three genes in this array, *ENA1*, *ENA2*, and *ENA5* but Sigma only has one gene (WIELAND *et al.* 1995). This difference in gene copy number is associated with a decreased salt tolerance in Sigma. Moreover, the Sigma genome contains a gene encoding the acetyltransferase *MPR1* which is not found in S288c (SHICHIRI *et al.* 2001) or most other yeast strains. This acetyltransferase was discovered because it confers resistance to the toxic proline analog L-azetidine-2-carboxylic acid (AZC) and has been subsequently shown to increase freeze tolerance and ethanol tolerance (DU and TAKAGI 2005; DU and TAKAGI 2007; SHICHIRI *et al.* 2001).

These prior studies have established that S288c and Sigma appear to be closely related but they have a few important differences at both the genomic and phenotypic level. However, there has been no attempt to systematically categorize these differences. Our study is the first to perform a genome-wide comparison of functions that differ between S288c and Sigma. This comparison was not performed by conventional linkage mapping but by building and screening

comparable deletion libraries for all the ORFs in the yeast genome. This analysis benefited from the ability to query every gene for a phenotype and allowed us rapidly to elucidate functional differences between the strains.

In the chapters that follow, I first present our findings regarding the characterization of a high quality assembly of the Sigma genome. We found that the nearly half of genes are completely identical between the strains, with most of the variation occurring at the subtelomeric regions. Using the genome sequence we were able to construct a Sigma deletion library that is comparable to the standard deletion library created in S288c. These libraries were screened for differences in essential genes and for adhesion defects. While relatively few genes show differences in essentiality, the adhesion phenotype shows a large amount of strain specific regualtion.

In Chapter Three of this thesis I present these findings of allele specific regulators of adhesion and a regulatory difference that is a consequence of the polymorphism in a specific gene. This work shows how the discovery of natural variation can be expedited by comparing deletion libraries, and offers a completely different approach than the traditional linkage studies in understanding the molecular mechanisms behind complex traits.

Figure 1-1 | **The Principle of linkage disequilibrium.** The causative mutation is indicated by a red triangle. Chromosomal stretches that are derived from the common ancestor of all mutant chromosomes are shown in light blue, whereas new stretches introduced by recombination are shown in dark blue. Over generations, markers that are physically close (that is, within the light-blue regions of present-day chromosomes) tend to remain associated with the ancestral mutation, even as recombination whittles down the region of association over time. (KRUGLYAK 2008).

Nature Reviews | Genetics
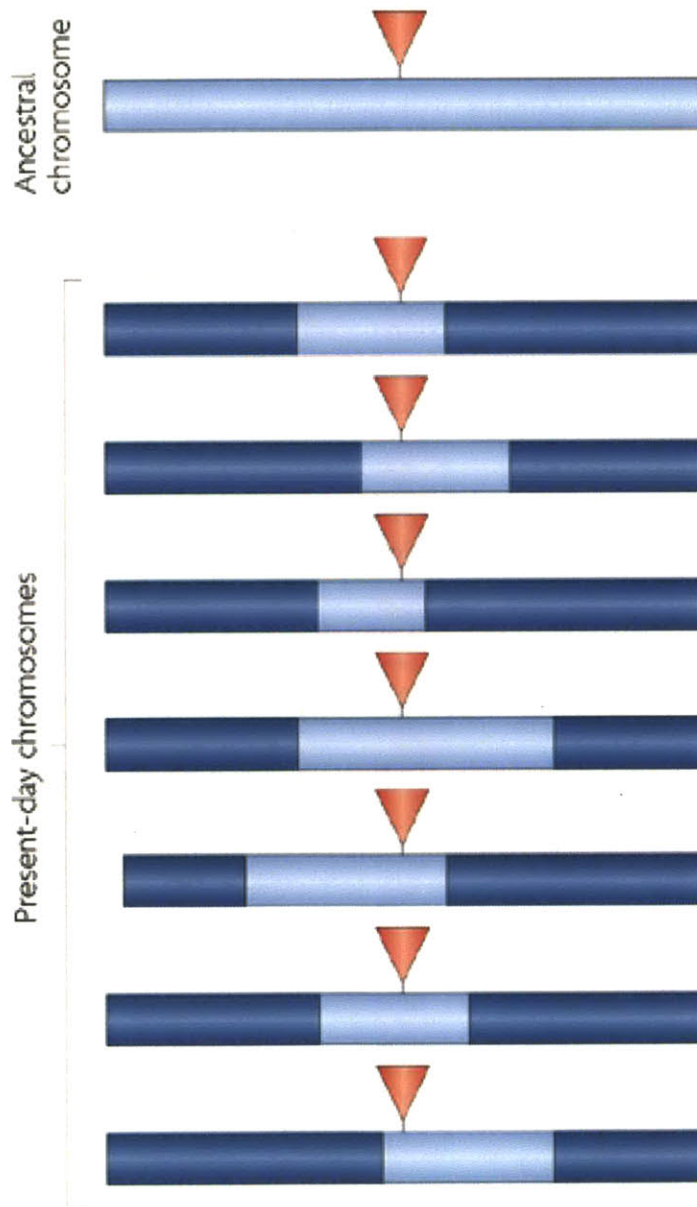
Figure 1-2 | **Example of transgressive segregation for a morphological trait.** Each cross represents one measurement of the length of the short axis of the mother cell. The first and second series are replicate cultures of BY and RM, respectively. Each cross of the last series represents the average value of one segregant measured in triplicate. Many segregant values fall outside the parental range, which illustrates the transgressive segregation of this trait. The inset drawing illustrates the trait definition. (NOGAMI *et al.* 2007)
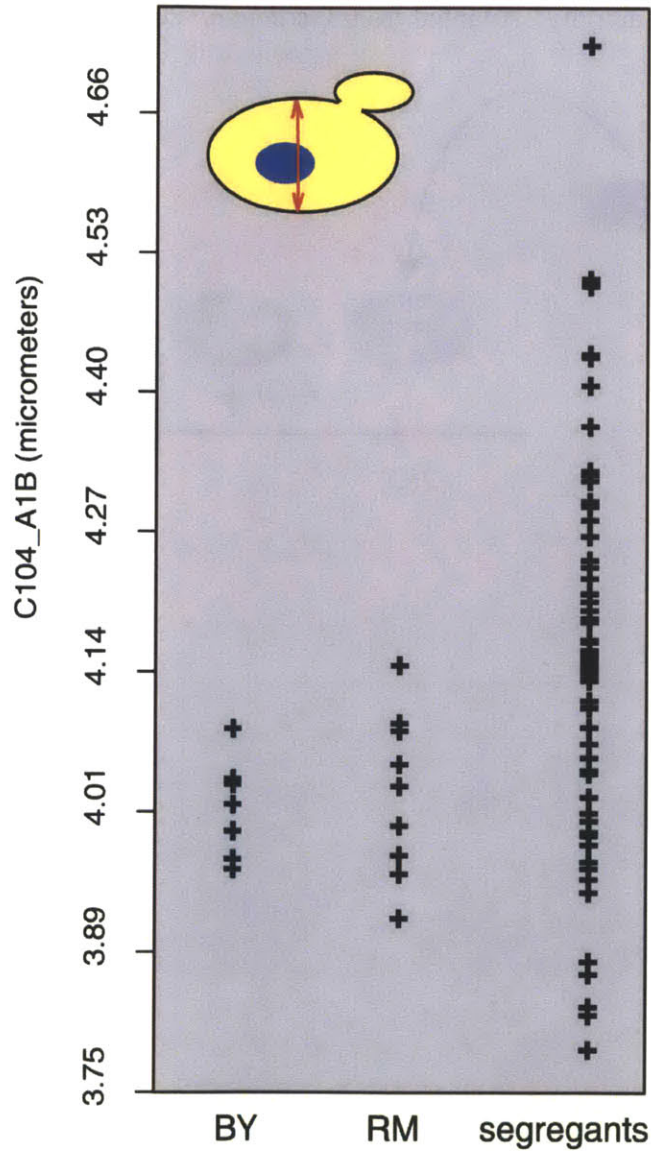
Figure 1-3 | **Molecular model of epistasis in sporulation efficiency.** Ime1p activity is required to proceed through meiosis, but at least four different naturally occurring polymorphisms influence Ime1p activity. Red X's represent causative variants affecting sporulation efficiency in the BC187 x YPS606 cross. The *RME1* allele 1(V) increases the amount of Rme1p and inhibits sporulaion. The *IME1* allele 2(O) is in the promoter of *IME1* and decreases the binding of Rme1p. The *IME1* allele 3(O) is in the coding sequence and increases Ime1p activity. Lastly, the RSF1 allele 4(O) is in the RSF1 coding sequence and increases its transcriptional activation activity. Adapted from Gertz *et al.* 2010.
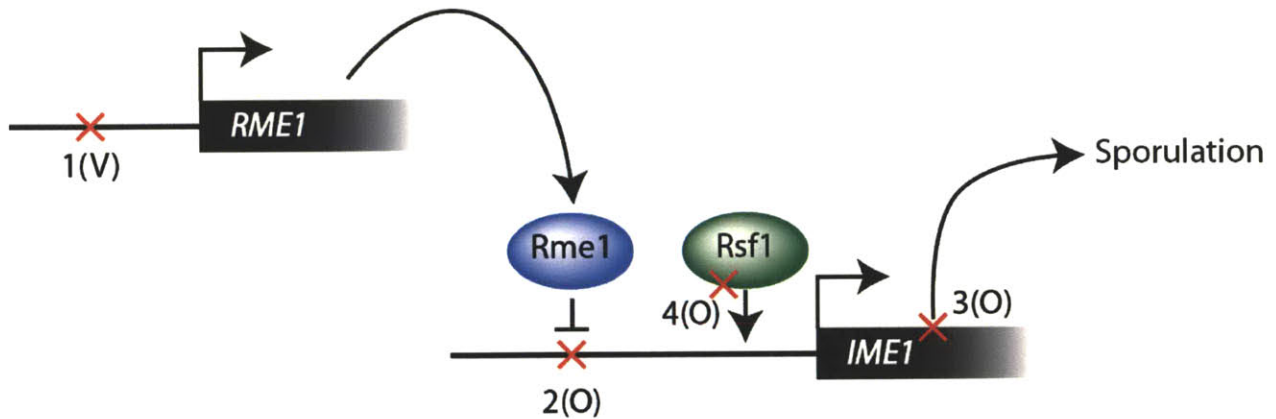
Figure 1-4 | **Multiple signaling pathways regulate *FLO11* expression.** The large promoter of *FLO11* contains interaction sites for many transcription factors that are regulated by multiple signaling pathways. This diagram shows a majority of the known regulators of *FLO11*. Shown in red are members of the PKA signaling pathway, and in blue are members of the fMAPK pathway. Shown in green are other genes that regulate *FLO11*, and whose regulation has been worked out to differing degrees.
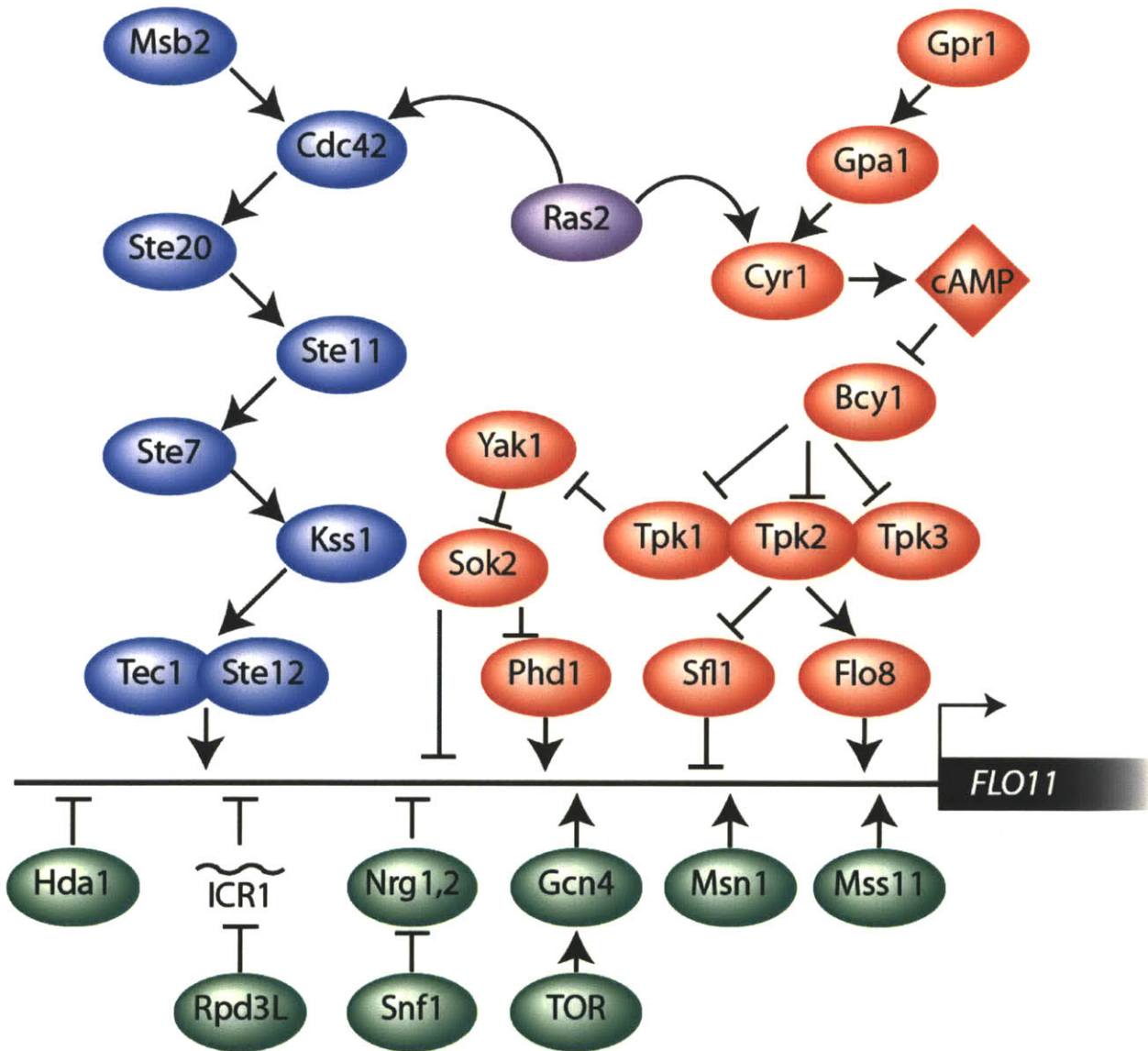
References

ALTSHULER, D., M. J. DALY and E. S. LANDER, 2008 Genetic mapping in human disease. Science **322**: 881-888.

BERGER, S. L., 2007 The complex language of chromatin regulation during transcription. Nature **447**: 407-412.

BOA, S., C. COERT and H. G. PATTERTON, 2003 Saccharomyces cerevisiae Set1p is a methyltransferase specific for lysine 4 of histone H3 and is required for efficient gene expression. Yeast **20**: 827-835.

BOER, V. M., S. AMINI and D. BOTSTEIN, 2008 Influence of genotype and nutrition on survival and metabolism of starving yeast. Proc Natl Acad Sci U S A **105**: 6930-6935.

BORNEMAN, A. R., J. A. LEIGH-BELL, H. YU, P. BERTONE, M. GERSTEIN *et al.*, 2006 Target hub proteins serve as master regulators of development in yeast. Genes Dev **20**: 435-448.

BRAUER, M. J., C. M. CHRISTIANSON, D. A. PAI and M. J. DUNHAM, 2006 Mapping novel traits by array-assisted bulk segregant analysis in Saccharomyces cerevisiae. Genetics **173**: 1813-1816.

BRAUS, G. H., O. GRUNDMANN, S. BRUCKNER and H. U. MOSCH, 2003 Amino acid starvation and Gcn4p regulate adhesive growth and FLO11 gene expression in Saccharomyces cerevisiae. Mol Biol Cell **14**: 4272-4284.

BREM, R. B., and L. KRUGLYAK, 2005 The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A **102**: 1572-1577.

BREM, R. B., G. YVERT, R. CLINTON and L. KRUGLYAK, 2002 Genetic dissection of transcriptional regulation in budding yeast. Science **296**: 752-755.

BRUCKNER, S., and H. U. MOSCH, 2012 Choosing the right lifestyle: adhesion and development in Saccharomyces cerevisiae. FEMS Microbiol Rev **36**: 25-58.

BUMGARNER, S. L., R. D. DOWELL, P. GRISAFI, D. K. GIFFORD and G. R. FINK, 2009 Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. Proc Natl Acad Sci U S A **106**: 18321-18326.

CARLBORG, O., and C. S. HALEY, 2004 Epistasis: too often neglected in complex trait studies? Nat Rev Genet **5**: 618-625.

CHOU, S., S. LANE and H. LIU, 2006 Regulation of mating and filamentation genes by two distinct Ste12 complexes in Saccharomyces cerevisiae. Mol Cell Biol **26**: 4794-4805.

COOK, J. G., L. BARDWELL, S. J. KRON and J. THORNER, 1996 Two novel targets of the MAP kinase Kss1 are negative regulators of invasive growth in the yeast Saccharomyces cerevisiae. Genes Dev **10**: 2831-2848.

COOK, J. G., L. BARDWELL and J. THORNER, 1997 Inhibitory and activating functions for MAPK Kss1 in the S. cerevisiae filamentous-growth signalling pathway. Nature **390**: 85-88.

CULLEN, P. J., and G. F. SPRAGUE, JR., 2012 The regulation of filamentous growth in yeast. Genetics **190**: 23-49.

DEMOGINES, A., E. SMITH, L. KRUGLYAK and E. ALANI, 2008 Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. PLoS Genet **4**: e1000123.

DIETVORST, J., and A. BRANDT, 2008 Flocculation in Saccharomyces cerevisiae is repressed by the COMPASS methylation complex during high-gravity fermentation. Yeast **25**: 891-901.

DU, X., and H. TAKAGI, 2005 N-acetyltransferase Mpr1 confers freeze tolerance on Saccharomyces cerevisiae by reducing reactive oxygen species. J Biochem **138**: 391-397.

DU, X., and H. TAKAGI, 2007 N-Acetyltransferase Mpr1 confers ethanol tolerance on Saccharomyces cerevisiae by reducing reactive oxygen species. Appl Microbiol Biotechnol **75**: 1343-1351.

EHRENREICH, I. M., N. TORABI, Y. JIA, J. KENT, S. MARTIS *et al.*, 2010 Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature **464**: 1039-1042.

FLEMING, A. B., and S. PENNINGS, 2001 Antagonistic remodelling by Swi-Snf and Tup1-Ssn6 of an extensive chromatin region forms the background for FLO1 gene regulation. EMBO J **20:** 5219-5231.

GATBONTON, T., M. IMBESI, M. NELSON, J. M. AKEY, D. M. RUDERFER *et al.*, 2006 Telomere length as a quantitative trait: genome-wide survey and genetic mapping of telomere length-control genes in yeast. PLoS Genet **2:** e35.

GERKE, J., K. LORENZ and B. COHEN, 2009 Genetic interactions between transcription factors cause natural variation in yeast. Science **323:** 498-501.

GERTZ, J., J. P. GERKE and B. A. COHEN, 2010 Epistasis in a quantitative trait captured by a molecular model of transcription factor interactions. Theor Popul Biol **77:** 1-5.

GIMENO, C. J., and G. R. FINK, 1994 Induction of pseudohyphal growth by overexpression of PHD1, a Saccharomyces cerevisiae gene related to transcriptional regulators of fungal development. Mol Cell Biol **14:** 2100-2112.

GIMENO, C. J., P. O. LJUNGDAHL, C. A. STYLES and G. R. FINK, 1992 Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. Cell **68:** 1077-1090.

GROUP, H. S. D. C. R., 1993 A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group. Cell **72:** 971-983.

GUO, B., C. A. STYLES, Q. FENG and G. R. FINK, 2000 A Saccharomyces gene family involved in invasive growth, cell-cell adhesion, and mating. Proc Natl Acad Sci U S A **97:** 12158-12163.

HALME, A., S. BUMGARNER, C. STYLES and G. R. FINK, 2004 Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. Cell **116:** 405-415.

IOANNIDIS, J. P., G. THOMAS and M. J. DALY, 2009 Validating, augmenting and refining genome-wide association signals. Nat Rev Genet **10:** 318-329.

JAKOBSDOTTIR, J., M. B. GORIN, Y. P. CONLEY, R. E. FERRELL and D. E. WEEKS, 2009 Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. PLoS Genet **5:** e1000337.

KEREM, B., J. M. ROMMENS, J. A. BUCHANAN, D. MARKIEWICZ, T. K. COX *et al.*, 1989 Identification of the cystic fibrosis gene: genetic analysis. Science **245:** 1073-1080.

KLIS, F. M., 1994 Review: cell wall assembly in yeast. Yeast **10:** 851-869.

KOBAYASHI, O., H. SUDA, T. OHTANI and H. SONE, 1996 Molecular cloning and analysis of the dominant flocculation gene FLO8 from Saccharomyces cerevisiae. Mol Gen Genet **251:** 707-715.

KOBAYASHI, O., H. YOSHIMOTO and H. SONE, 1999 Analysis of the genes activated by the FLO8 gene in Saccharomyces cerevisiae. Curr Genet **36:** 256-261.

KOHLER, T., S. WESCHE, N. TAHERI, G. H. BRAUS and H. U. MOSCH, 2002 Dual role of the Saccharomyces cerevisiae TEA/ATTS family transcription factor Tec1p in regulation of gene expression and cellular development. Eukaryot Cell **1:** 673-686.

KROGAN, N. J., J. DOVER, S. KHORRAMI, J. F. GREENBLATT, J. SCHNEIDER *et al.*, 2002 COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. J Biol Chem **277:** 10753-10755.

KRON, S. J., C. A. STYLES and G. R. FINK, 1994 Symmetric cell division in pseudohyphae of the yeast Saccharomyces cerevisiae. Mol Biol Cell **5:** 1003-1022.

KROYMANN, J., and T. MITCHELL-OLDS, 2005 Epistasis and balanced polymorphism influencing complex trait variation. Nature **435:** 95-98.

KRUGLYAK, L., 2008 The road to genome-wide association studies. Nat Rev Genet **9:** 314-318.

KUCHIN, S., V. K. VYAS and M. CARLSON, 2002 Snf1 protein kinase and the repressors Nrg1 and Nrg2 regulate FLO11, haploid invasive growth, and diploid pseudohyphal differentiation. Mol Cell Biol **22:** 3994-4000.

LAIZE, V., F. TACNET, P. RIPOCHE and S. HOHMANN, 2000 Polymorphism of Saccharomyces cerevisiae aquaporins. Yeast **16:** 897-903.

LIU, H., C. A. STYLES and G. R. FINK, 1993 Elements of the yeast pheromone response pathway required for filamentous growth of diploids. Science **262:** 1741-1744.

LIU, H., C. A. STYLES and G. R. FINK, 1996 Saccharomyces cerevisiae S288C has a mutation in FLO8, a gene required for filamentous growth. Genetics **144:** 967-978.

LO, W. S., and A. M. DRANGINIS, 1996 FLO11, a yeast gene related to the STA genes, encodes a novel cell surface flocculin. J Bacteriol **178:** 7144-7151.

LO, W. S., and A. M. DRANGINIS, 1998 The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by Saccharomyces cerevisiae. Mol Biol Cell **9:** 161-171.

MADHANI, H. D., and G. R. FINK, 1997 Combinatorial control required for the specificity of yeast MAPK signaling. Science **275:** 1314-1317.

MALCHER, M., S. SCHLADEBECK and H. U. MOSCH, 2011 The Yak1 protein kinase lies at the center of a regulatory cascade affecting adhesive growth and stress resistance in Saccharomyces cerevisiae. Genetics **187:** 717-730.

MANOLIO, T. A., F. S. COLLINS, N. J. COX, D. B. GOLDSTEIN, L. A. HINDORFF *et al.*, 2009 Finding the missing heritability of complex diseases. Nature **461:** 747-753.

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

NOGAMI, S., Y. OHYA and G. YVERT, 2007 Genetic complexity and quantitative trait loci mapping of yeast morphological traits. PLoS Genet **3:** e31.

O'ROURKE, S. M., and I. HERSKOWITZ, 1998 The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in Saccharomyces cerevisiae. Genes Dev **12:** 2874-2886.

OCTAVIO, L. M., K. GEDEON and N. MAHESHRI, 2009 Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. PLoS Genet **5:** e1000673.

PAN, X., and J. HEITMAN, 1999 Cyclic AMP-dependent protein kinase regulates pseudohyphal differentiation in Saccharomyces cerevisiae. Mol Cell Biol **19:** 4874-4887.

PERLSTEIN, E. O., D. M. RUDERFER, G. RAMACHANDRAN, S. J. HAGGARTY, L. KRUGLYAK *et al.*, 2006 Revealing complex traits with small molecules and naturally recombinant yeast strains. Chem Biol **13:** 319-327.

PITONIAK, A., B. BIRKAYA, H. M. DIONNE, N. VADAIE and P. J. CULLEN, 2009 The signaling mucins Msb2 and Hkr1 differentially regulate the filamentation mitogen-activated protein kinase pathway and contribute to a multimodal response. Mol Biol Cell **20:** 3101-3114.

REYNOLDS, T. B., and G. R. FINK, 2001 Bakers' yeast, a model for fungal biofilm formation. Science **291:** 878-881.

RIORDAN, J. R., J. M. ROMMENS, B. KEREM, N. ALON, R. ROZMAHEL *et al.*, 1989 Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science **245:** 1066-1073.

ROBERTS, R. L., and G. R. FINK, 1994 Elements of a single MAP kinase cascade in Saccharomyces cerevisiae mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev **8:** 2974-2985.

ROMMENS, J. M., M. C. IANNUZZI, B. KEREM, M. L. DRUMM, G. MELMER *et al.*, 1989 Identification of the cystic fibrosis gene: chromosome walking and jumping. Science **245:** 1059-1065.

RONALD, J., R. B. BREM, J. WHITTLE and L. KRUGLYAK, 2005 Local regulatory variation in Saccharomyces cerevisiae. PLoS Genet **1:** e25.

RYTKA, J., 1975 Positive selection of general amino acid permease mutants in Saccharomyces cerevisiae. J Bacteriol **121:** 562-570.

SCHACHERER, J., J. A. SHAPIRO, D. M. RUDERFER and L. KRUGLYAK, 2009 Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. Nature **458:** 342-345.

SEGRE, A. V., A. W. MURRAY and J. Y. LEU, 2006 High-resolution mutation mapping reveals parallel experimental evolution in yeast. PLoS Biol **4:** e256.

SHICHIRI, M., C. HOSHIKAWA, S. NAKAMORI and H. TAKAGI, 2001 A novel acetyltransferase found in Saccharomyces cerevisiae Sigma1278b that detoxifies a proline analogue, azetidine-2-carboxylic acid. J Biol Chem **276:** 41998-42002.

SMITS, G. J., J. C. KAPTEYN, H. VAN DEN ENDE and F. M. KLIS, 1999 Cell wall dynamics in yeast. Curr Opin Microbiol **2:** 348-352.

SMUKALLA, S., M. CALDARA, N. POCHET, A. BEAUVAIS, S. GUADAGNINI et al., 2008 FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. Cell **135:** 726-737.

STRATFORD, M., and S. ASSINDER, 1991 Yeast flocculation: Flo1 and NewFlo phenotypes and receptor structure. Yeast **7:** 559-574.

TANGHE, A., P. VAN DIJCK, F. DUMORTIER, A. TEUNISSEN, S. HOHMANN et al., 2002 Aquaporin expression correlates with freeze tolerance in baker's yeast, and overexpression improves freeze tolerance in industrial strains. Appl Environ Microbiol **68:** 5981-5989.

TONG, A. H., and C. BOONE, 2006 Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol Biol **313:** 171-192.

TORABI, N., and L. KRUGLYAK, 2011 Variants in SUP45 and TRM10 underlie natural variation in translation termination efficiency in Saccharomyces cerevisiae. PLoS Genet **7:** e1002211.

VERSTREPEN, K. J., A. JANSEN, F. LEWITTER and G. R. FINK, 2005 Intragenic tandem repeats generate functional variability. Nat Genet **37:** 986-990.

VERSTREPEN, K. J., and F. M. KLIS, 2006 Flocculation, adhesion and biofilm formation in yeasts. Mol Microbiol **60:** 5-15.

VINEIS, P., and N. PEARCE, 2011 Genome-wide association studies may be misinterpreted: genes versus heritability. Carcinogenesis **32:** 1295-1298.

VINOD, P. K., N. SENGUPTA, P. J. BHAT and K. V. VENKATESH, 2008 Integration of global signaling pathways, cAMP-PKA, MAPK and TOR in the regulation of FLO11. PLoS One **3:** e1663.

VYAS, V. K., S. KUCHIN, C. D. BERKEY and M. CARLSON, 2003 Snf1 kinases with different beta-subunit isoforms play distinct roles in regulating haploid invasive growth. Mol Cell Biol **23:** 1341-1348.

WEEDON, M. N., H. LANGO, C. M. LINDGREN, C. WALLACE, D. M. EVANS et al., 2008 Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet **40:** 575-583.

WIELAND, J., A. M. NITSCHE, J. STRAYLE, H. STEINER and H. K. RUDOLPH, 1995 The PMR2 gene cluster encodes functionally distinct isoforms of a putative Na+ pump in the yeast plasma membrane. EMBO J **14:** 3870-3882.

WINZELER, E. A., C. I. CASTILLO-DAVIS, G. OSHIRO, D. LIANG, D. R. RICHARDS et al., 2003 Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics **163:** 79-89.

YVERT, G., R. B. BREM, J. WHITTLE, J. M. AKEY, E. FOSS et al., 2003 Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. Nat Genet **35:** 57-64.

ZEITLINGER, J., I. SIMON, C. T. HARBISON, N. M. HANNETT, T. L. VOLKERT et al., 2003 Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. Cell **113:** 395-404.

Chapter 2

Genotype to Phenotype: A Complex Problem

Author's Note – The material presented in this chapter was originally published as: R. D. Dowell, O. Ryan, A. Jansen, D. Cheung, S. Agarwala, T. Danford, D. A. Bernstein, P. A. Rolfe, L. E. Heisler, **B. Chin**, C. Nislow, G. Giaever, P. C. Phillips, G. R. Fink, D. K. Gifford and C. Boone (2010) "Genoype to Phenotype: A Complex Problem" Science 328, 469.

I contributed to this work at a number of different points. Initially I helped to decipher potentially incorrectly assembled regions of the Sigma genome sequence. This was done both computationally and experimentally by performing PCR and CHEF gel analysis of regions of the genome to validate insertions and rearrangements. I later assisted with the analysis of the Sigma specific essential phenotypes of *SKI7* and *BEM1*. Much of the data was moved into supplementary information, and the relevant sections are provided here.
The deletion library and genome sequence presented here provided the foundation for the comparisons presented in the later chapters.

Rapid genome-sequencing methods coupled with whole-genome transcription pro- filing suggests that it may be possible to predict phenotype from a genotype. Human genetic association studies of common single- nucleotide polymorphisms (SNPs) explain only a fraction of phenotypic variation among individuals (DICKSON *et al.* 2010). This may be due to rare SNPs (Dickson et al. 2010), structural (Korbel et al. 2007) and epigenetic variants, or multiple alleles with additive effects or synergistic genetic interactions associated with complex combinations of genetic variation (Hartman et al. 2001).

To address the genotype-to-phenotype problem, we developed a simple comparative model for the budding yeast Saccharomyces cerevisiae that enables a comprehensive assessment of the genetic mechanisms leading to different phenotypes for the same mutation in two different genetic backgrounds. The strain Σ1278b mates and forms viable meiotic progeny with the

reference strain, S288c, and the divergence between the two strains is roughly equal to the divergence between the genomes of two humans (Wang et al. 2008).

We sequenced and assembled the 12-Mb Σ1278b genome, annotating 6923 open reading frames (ORFs) and RNAs, of which 6848 have orthologs within S288c. The order of genes between the strains was the same (except in the highly variable subtelomeric regions), and the sequence of 46% of the Σ1278b ORFs was identical to those in S288c. Differences between the strains were largely due to small insertions and deletions or SNPs, with an average SNP density of 3.2 per kilobase.

We deleted ~5100 genes within Σ1278b to systematically compare identical deletion mutants (Giaever et al. 2002). In particular, we identified "conditional essentials," those genes required uniquely for viability in either strain (Table S2-1). We scored colonies as dead or alive and surveyed all vital pathways for individual-specific genetic interactions. We expected such conditional essential genes to be rare because the genomes of Σ1278b and S288c are nearly identical.

Although 894 genes were essential in both S288c and Σ1278b, 44 genes were essential only in Σ1278b and 13 genes were essential only in S288c (Figure 2-1A). The conditional subsets included genes of various functions; however, the Σ1278b subset was enriched for genes involved in mRNA metabolic process, whereas the S288c set was enriched for genes annotated to SRP-dependent cotranslational targeting. These biological biases suggest that these phenotypes result from genetic interactions associated with an individual genotype.

Hybrid strain crosses and tetrad analysis focusing on 18 mutants that were lethal in Σ1278b with wild-type levels of fitness when deleted in S288c were used to investigate

conditional essentiality. We mated viable haploid S288c deletion mutants to wild-type Σ1278b

and analyzed the hybrid diploid progeny by tetrad analysis. The number of viable meiotic

progeny carrying the deletion allele is related to the number of unlinked background- specific

modifiers that contribute to the genetic interaction. In all 18 cases, the conditional phenotype was

associated with numerous modifier genes that differ between strains. The simplest cases, *SKI7*

and *BEM1*, are likely due to a genetic interaction with at least two or more modifiers, but all

other cases were more complex (Figure 2-1B). Thus, our analysis showed that conditional

essentiality is almost always a consequence of complex genetic interactions involving multiple

modifiers associated with strain-specific genetic variation rather than classic digenic synthetic

lethality (Costanzo et al. 2010; Giaever et al. 2002).

Our genome-wide survey of conditionally essential genes demonstrates that in most cases

a complex set of background-specific modifiers influence a mutation whose phenotype differs

between individuals. These results raise the possibility that similar complex modifiers may

largely explain the difficulty in identifying the genetic basis for individual phenotypes. The

potential for genetic interactions to control individual phenotypes becomes even more important

if different combinations of alleles can lead to the same physiological state. The ability to

identify these conditional essential phenotypes in yeast provides a framework to unravel the

fundamental principles of genetic networks resulting from natural variation, including those that

underlie human disease.

Figure 2-1 | (A) Most S288c essential genes are also essential in Σ1278b (94%); however, ~5% are essential only in the Σ 1278b genetic background, whereas ~1% are essential only in S288c. (B) Conditional essential genes in Σ 1278b are the consequence of complex genetics. c2 tests indicated the number of modifiers associated with conditional essentiality.

A **Σ1278b: 44 Genes (5%)**
Conditional Essential

894 Genes (94%)
Unconditional
Essential

S288c: 13 Genes (1%)
Conditional Essential

B **Number of Modifiers**

■ Two genes
■ Three genes
■ Four or more genes

BEM1 SKI7 LEA1 FMP27 YPR015C SBH2 TMA108 PEP7 ZWF1 MTO1 SKI8 YKR075C YHR009C AAT2 YDL089W PEP12 VPS16 MCM22
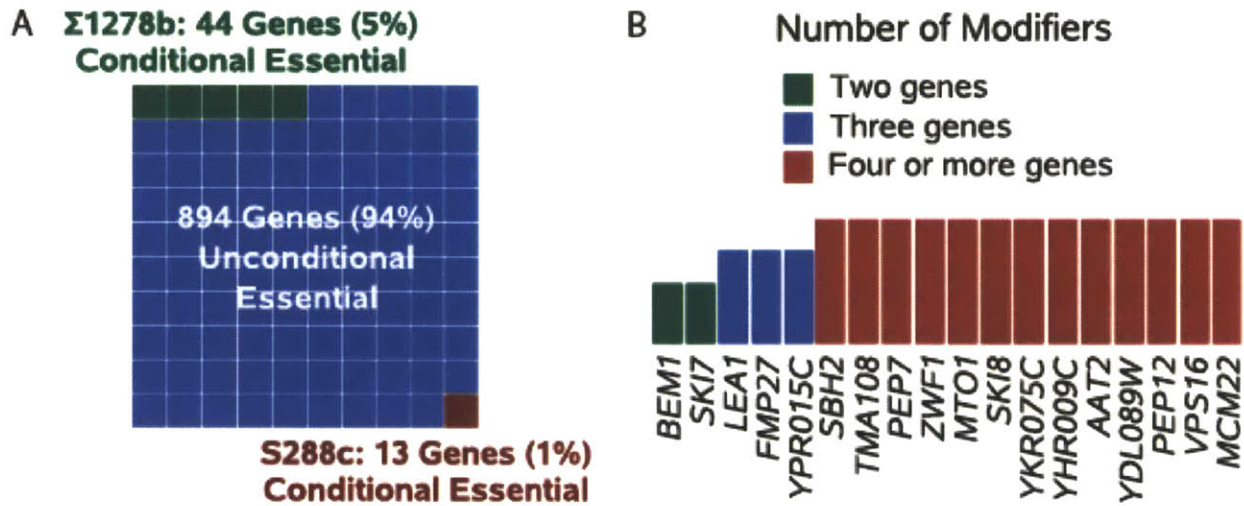
References and notes

COSTANZO, M., A. BARYSHNIKOVA, J. BELLAY, Y. KIM, E. D. SPEAR *et al.*, 2010 The genetic landscape of a cell. Science **327:** 425-431.

DICKSON, S. P., K. WANG, I. KRANTZ, H. HAKONARSON and D. B. GOLDSTEIN, 2010 Rare variants create synthetic genome-wide associations. PLoS Biol **8:** e1000294.

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

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

KORBEL, J. O., A. E. URBAN, J. P. AFFOURTIT, B. GODWIN, F. GRUBERT *et al.*, 2007 Paired-end mapping reveals extensive structural variation in the human genome. Science **318:** 420-426.

WANG, J., W. WANG, R. LI, Y. LI, G. TIAN *et al.*, 2008 The diploid genome sequence of an Asian individual. Nature **456:** 60-65.

Genotype to Phenotype: A Complex Problem
Dowell, Ryan, et. al.

Supporting Online Material

Data files S1 and S2 are available at
http://mcdb.colorado.edu/labs1/dowelllab/pubs/DowellRyan/

Genome

Yeast strains used are listed in Table S2-2. Yeast cultures were grown as described (S1 (SHERMAN *et al.* 1986)). Real-time PCR utilized the ABI 7500 system (Applied Biosystems, Foster City, CA) and was carried out with the appropriate enzymes and chemicals from Applied Biosystems as recommended by the supplier. CHEF chromosome separation was performed with a BioRad CHEF-DRII (BioRad, Hercules, CA) with the protocol supplied by BioRad.

Sequence and Assembly   We produced whole genome shotgun sequence from two plasmid libraries (4kb and 10kb inserts) of the Saccharomyces cerevisiae strain Σ1278b, sub-strain 10560-6B. Genomic DNA was isolated with the Qiagen Genomic-tip kit (Qiagen, Valencia, CA) following the manufacturers' protocol. Initial sequence was generated with the whole genome sequencing and assembly methodology utilized to sequence the RM11-1a strain (BROAD). The resulting 7.3X Arachne long read assembly contained 12.2 Mb in 111636 sequence reads, 357 contigs and 51 scaffolds.

In addition 20 million 36 nucleotide reads were generated using an Illumina Genome Analyzer located at the Whitehead Institute Genome Technology Core. Samples for Illumina sequencing were purified with the standard protocols outlined in their genomic DNA sample prep kit (Illumina, San Diego, CA). Three lanes of cluster generations were performed on an Illumina cluster station with 2pM sequencing libraries for each lane. These reads were assembled with Velvet (ZERBINO and BIRNEY 2008) (v0.6.03) with a coverage cutoff of 5 and a minimum length of 100 nts, resulting in 11.3 Mb in 5419 contigs.

The BlastZ (v7) (BLANCHETTE *et al.* 2004) and MUMer (v3.19) (DELCHER *et al.* 2002) software packages were utilized to align the long read scaffolds to the S288c chromosomes

[Saccharomyces Genome Database (SGD) March 2009; http://www.yeastgenome.org/].

Contour-clamped homogeneous electric field (CHEF) gel electrophoresis and site specific PCR were used to correct the misassembly of four scaffolds and ascertain their location, size, and boundaries. In addition, one scaffold had no clear S288c correspondence and was localized in Σ1278b by CHEF gel. The short read scaffolds were then utilized to fill in gaps and correct poor quality segments within the chromosomes with a combination of BLAT (Nov 2006 (KENT 2002)), fsa (v1.07; (BRADLEY et al. 2009)), and manual inspection.

Annotation   We used three methods to identify potential ORFs in the Σ1278b sequence: (i) directly mapping S288c ORFs, (ii) identification of long open reading frames, and (iii) the genefinder GlimmerHMM (MAJOROS et al. 2004; SALZBERG et al. 1999). The S288c ORFs were mapped to Σ1278b by identifying the best BLAT (Nov 2006 (KENT 2002)) hit utilizing the complete set of ORFs obtained from the Saccharomyces database (SGD: http://www.yeastgenome.org/; March 2009). GlimmerHMM (MAJOROS et al. 2004; SALZBERG et al. 1999)) was trained on the non-mitochondrial S288c ORFs.

We identified the S288c orthologs within the Σ1278b sequence by a combination of sequence identity and appropriate synteny (KELLIS et al. 2003). The remaining potential Σ1278b ORFs were compared to the non-redundant database (NCBI May 2008) by WU-Blast (v2.0; http://blast.wustl.edu/) to identify previously characterized genes not present in S288c. The gene names for the Σ1278b genes with S288c homologs were annotated according to their S288c counterpart. The annotation of Σ1278b genes absent from S288c is from a comparison to the non-redundant database (NCBI May 2008). Functional annotations, in particular GO associations, were taken from the S288c counterpart.

Noncoding RNAs were annotated by a combination of methods. tRNAscanSE v1.23 (LOWE and EDDY 1997) identified tRNAs within the Σ1278b genome. Other RNAs features were identified by BLAT (Nov 2006 (KENT 2002)) from the S288c counterpart, taking into consideration synteny with surrounding ORF annotations.

The majority of differences between S288c and Σ1278b excluding subtelomeric regions, were single nucleotide polymorphisms (SNPs) and small insertions or deletions (indels) distributed throughout the chromosomes. The Σ1278b strain has an average SNP density of 3.2 SNPs per kilobase, as determined by alignments generated by fsa (v1.07; (BRADLEY et al. 2009)). Sequence comparison did not uncover any obvious duplication of genes essential in S288c.

Deletion Library Construction

Deletion cassettes were PCR amplified such that they were flanked by 100-250 base pairs of S288c homology for each casette. Primers were designed with Primer3 software (ROZEN and SKALETSKY 2000) with parameters set between 100-300bp beyond the START and STOP codons of each S288c open reading frame, with comparable melting temperatures and GC content.

Deletion cassettes were colony PCR amplified with Hi Fidelity Enzyme (Roche, Nutley, NJ) for 40 cycles. Each deletion casette contains the kanamycin (KanMX) marker flanked by molecular barcodes and their common primers (GIAEVER et al. 2002) and thus each deletion allele and its corresponding molecular barcodes were transferred from the S288c deletion mutant collection to the Σ1278b deletion mutant collection.

Fourty-seven (47) Σ1278b specific genes were deleted from the Σ1278b genome. Primers

were designed on the basis of the Σ1278b genome sequence, with 50 base pair 5' tails of

homology to the upstream and downstream of each specific gene to be deleted. Two unique

molecular barcodes were assigned to each deletion mutant. PCR products (KanMX cassette +

homologous DNA) were transformed with lithium acetate based transformation into strain

YSWT3.

Transformants were recovered for 4 hours in YEPD liquid and then plated onto YEPD

containing 200mg/ml G418. Colonies derived from a single transformation event were colony

PCR confirmed by with primers that lie >350 base pairs upstream from START (KanMX

internal primer sequence 5'-TCTGCAGCGAGGAGCCGTAAT-3').

Identifying Essential Genes by Random Spore Analysis (RSA) Haploid mutant strains

were isolated by sporulating the diploid heterozygous deletion mutants for 4 days on solid

sporulation medium. MATa meiotic progeny were germinated on haploid selection medium, SD-

HIS/ARG/LYS+cananvanine+thialysine (TONG et al. 2004), which is minimal medium lacking

histidine, arginine and lysine but included the toxic amino acid analogs canavanine and

thialysine, which provides a counter-selection against heterozygous diploids. The lack of

histidine selects for cells expressing STE2pr-sphis5, a construct that places the S. pombe his5

gene under the control of the MATaspecific STE2 promoter. Following their germination,

essential genes were identified by replica plating the haploid meiotic progeny from haploid

selection medium to YEPD+G418, which selects for growth of haploid deletion mutant cells.

Essential gene function was identified by the absence of viable colonies on the YEPD+G418

plate.

Tetrad Confirmation   For genes determined to be essential for viability by RSA in the Σ1278b background but non-essential for viability in S288c background (as defined by SGD), or genes determined to be non-essential for viability in the Σ1278b background but essential for viability in the S288c background, tetrad analysis was performed. Heterozygous diploid mutant strains of both backgrounds, S288c and Σ1278b, were sporulated on solid sporulation medium for 1-2 weeks. Asci were digested with Zymolyase and tetrads were dissected onto YEPD and grown 4 days at 30°C. Plates were photographed and replica plated on YEPD+G418 to follow the segregation patterns of knockout alleles relative to fitness phenotype (See Table S2-1). GO enrichments were calculated using SGD's GO Term Finder.

Hybrid S288c/Σ1278b Tetrad Dissection   A hybrid wild-type diploid strain (Y12868) was created crossing S288c MATa (Y1239) to Σ1278b MATα (Y3295) and zygotes were isolated with a tetrad dissecting microscope. To determine the naturally occurring synthetic lethality rate between wild type S288c and wild type Σ1278b, 129 tetrads were dissected identifying 504 meiotic segregants, of which 6 failed to germinate (1.19% lethality). Hybrid mutant strains were created by crossing the MATa deletion mutants from the S288c collection to Y3295. Diploids were sporulated for 5 days and tetrads dissected on YEPD plates. Tetrad segregation pattern were tabulated for each hybrid deletion mutant. A chi-squared statistic ($\chi2$) was then utilized to test three separate hypothesis: (1) a single unlinked modifier explains the inheritance patterns (1:1:4 ratio expected); (2) three unlinked modifiers explain the inheritance patterns; and (3) complex genetics (many loci) make the inheritance patterns indistinguishable from empirically observed background, from the wild type vs wild type cross (Y1239 diploid). In all cases, a p-value was calculated for the $\chi2$ statistic using Microsoft Excel's CHIDIST function (Table S2-3).

Theoretical population genetics suggests that loss-of-function mutations are predicted to accumulate to high levels in a population for genes with a single and closely linked synthetic lethal partner because linkage prevents clearance of these mutations through mating and meiotic recombination (PHILLIPS 1998). To test for the possibility that the segregation patterns we observed are caused by the tight linkage of a conditional essential gene to a single second gene that causes its lethal phenotype, we examined two conditional essential genes in greater detail. We transformed Y12868 with either *mto1*Δ::KanMX or *pep12*Δ::KanMX and then sporulated the resultant heterozygous deletion mutant. Sequencing proximal to the integrated deletion cassette allowed us to determine the parental locus (S288c or Σ1278b) into which the deletion cassette integrated. If a synthetic lethal partner present only in Σ1278b were tightly linked to the conditionally essential gene, equivalent to having a single tightly linked suppressor in the S288c genetic background, then the deletion allele integrated into the Σ1278b chromosome should yield only inviable progeny, whereas the deletion allele integrated into the S288c chromosome should yield only viable spores. Between 60 and 70 tetrads were dissected for each mutant and segregation patterns of lethality and G418-resistance were scored. For both MTO1 and PEP12, we found that the deletion alleles integrated into both S288c and Σ1278b generated relatively few inviable spores. These data, along with the segregation of inviability, show that conditional essentiality is most often a consequence of complex synthetic lethality.

Supplemental Figures:

Figure S2-1 | **Histogram showing nucleotide differences between ORFs in S288c and Σ1278b.** Of those genes less than 90% identical, nearly one half are contained within subtelomeric regions. The percent identity labels indicate the lower range of the bar, with the 100 % bin containing only those genes that are absolutely identical between the strains (no SNPs or indels). Bins are chosen to emphasize the fact that 94% of all genes are 99% identical or better. Genes containing N's in the Σ1278b genome are excluded. Pairwise percent identity is calculated as the number of identical nucleotides divided by length of the shorter sequence on alignments generated by ClustalW (v1.83 (THOMPSON *et al.* 1994)).

Histogram of SNP changes
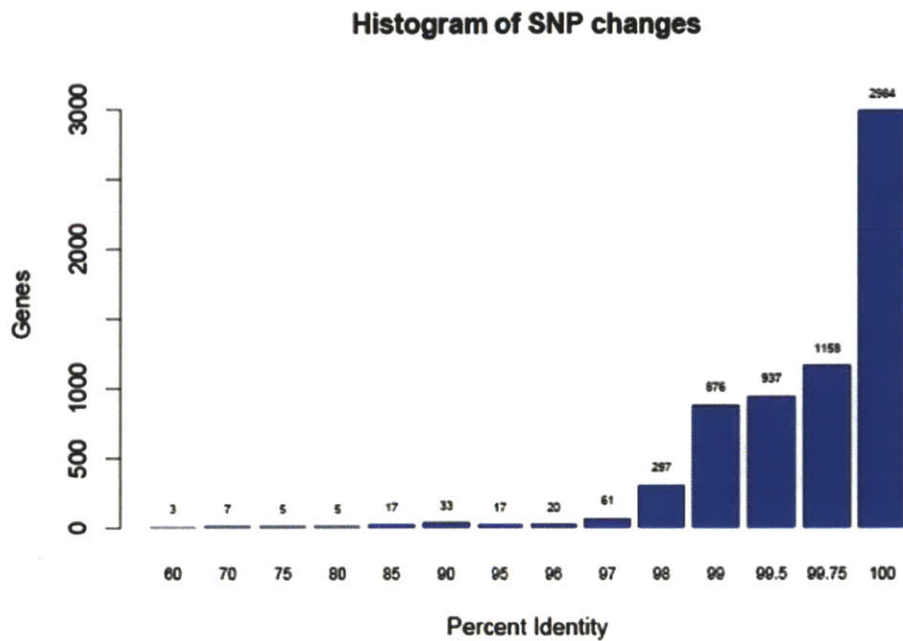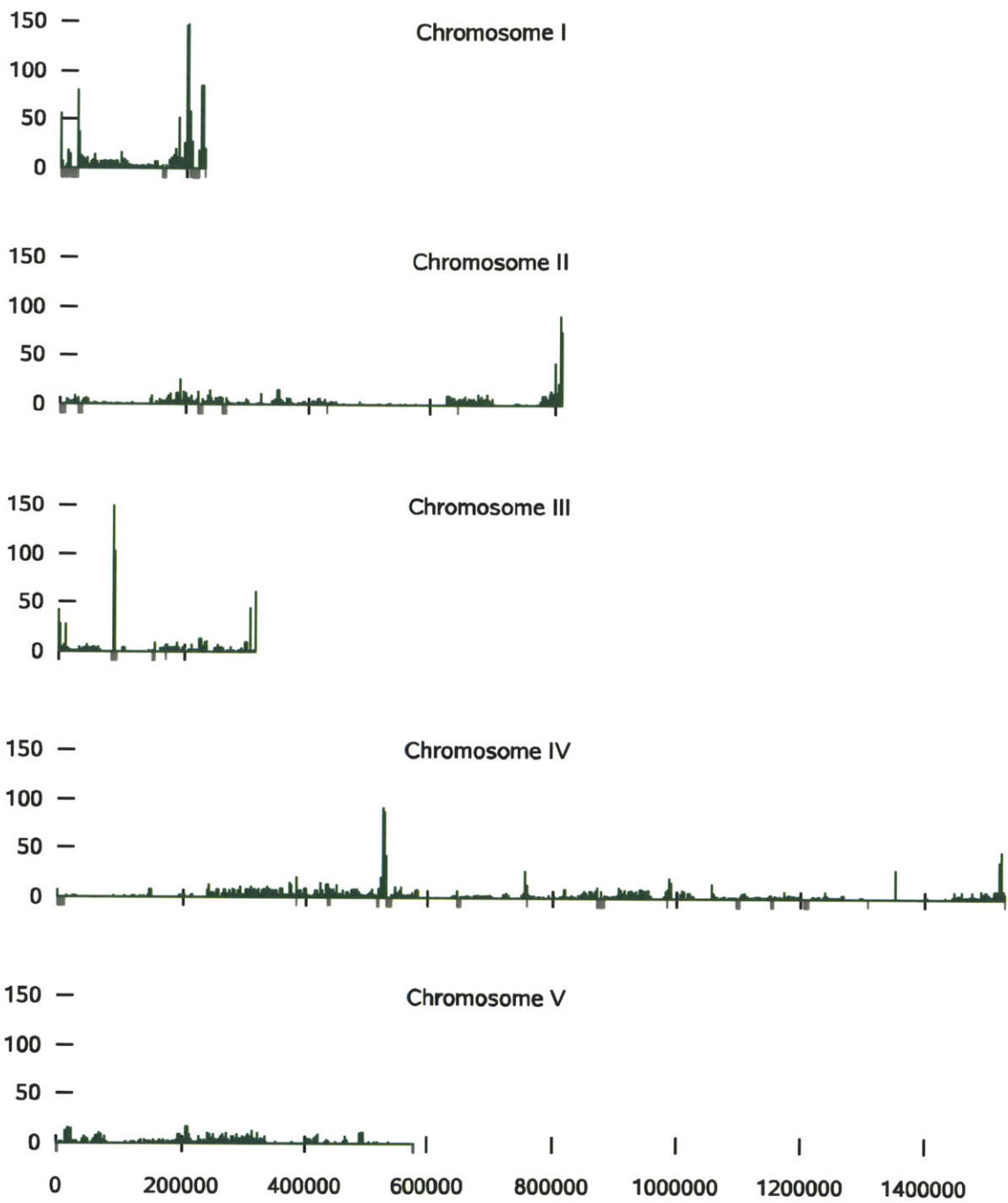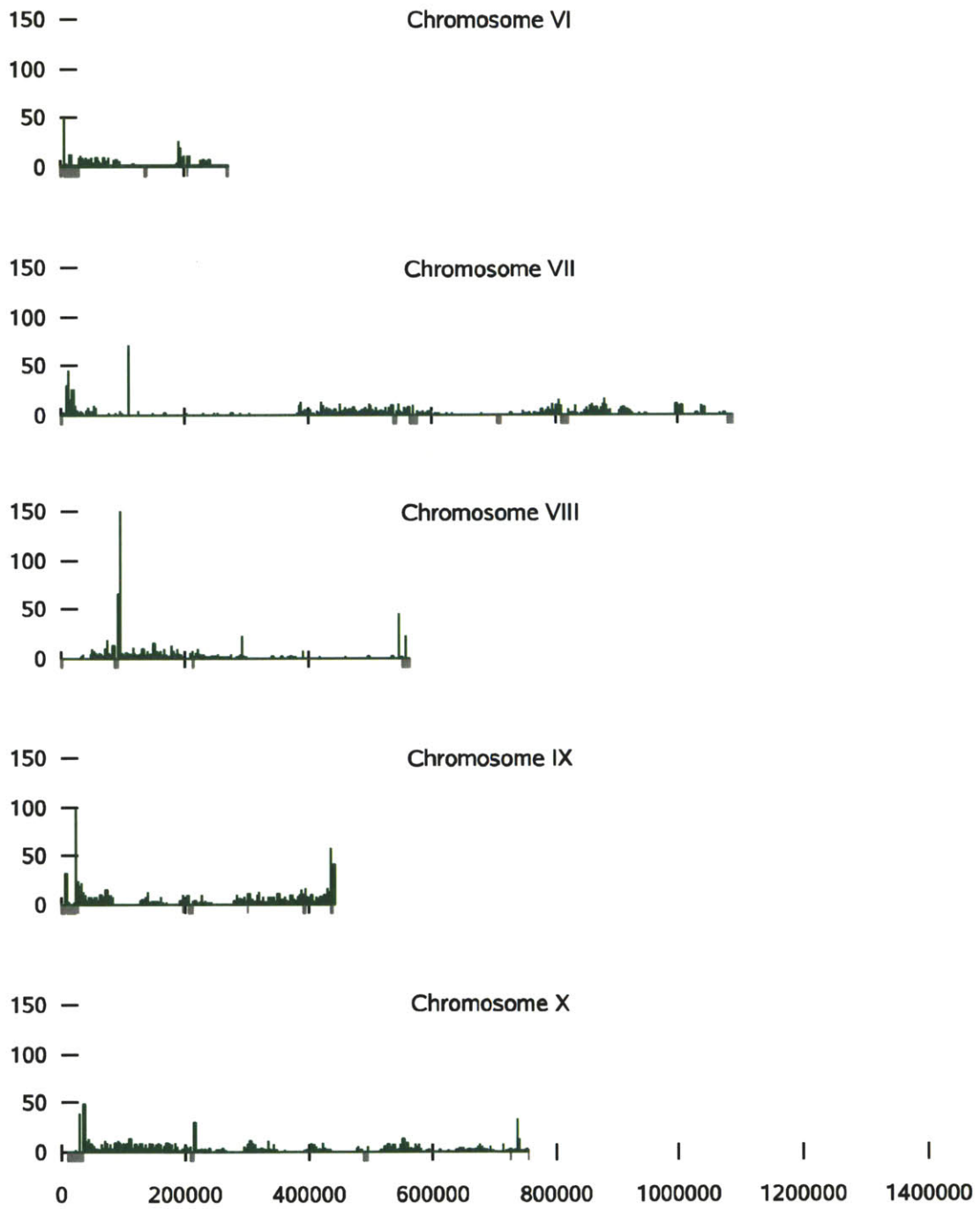
Figure S2: Continued next page
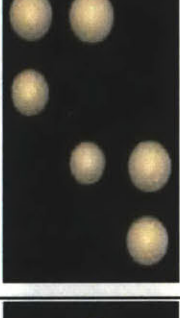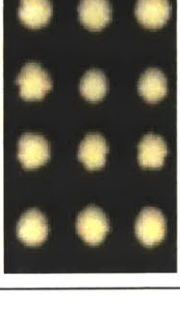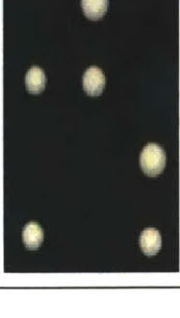
Figure S2: Continued next page

48
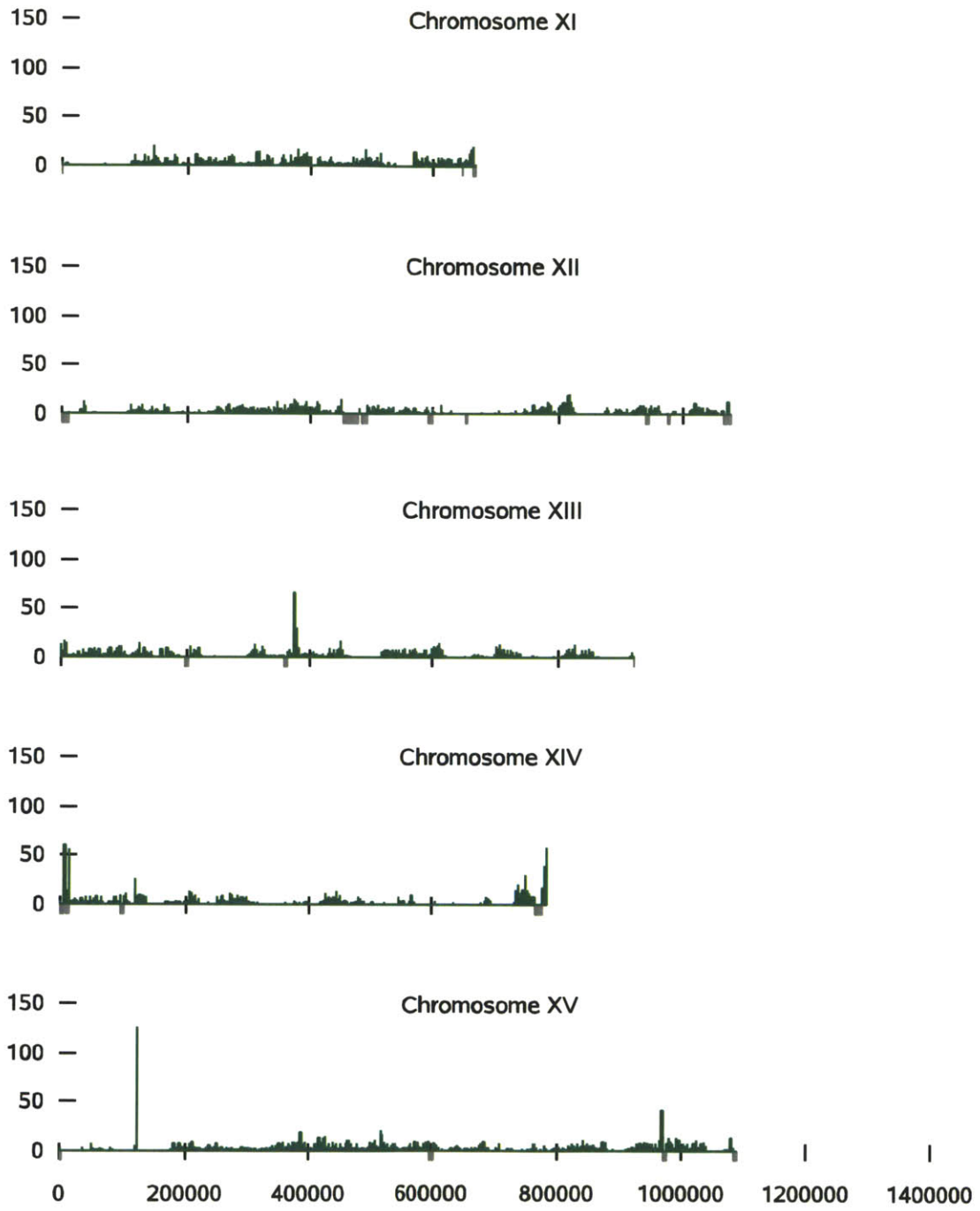
Figure S2: Continued next page
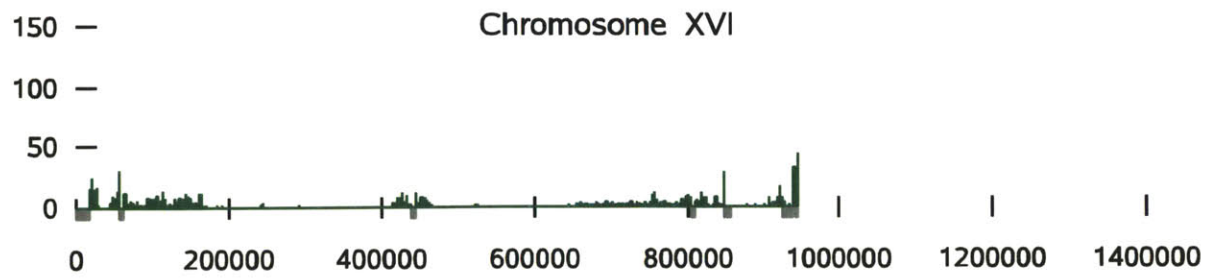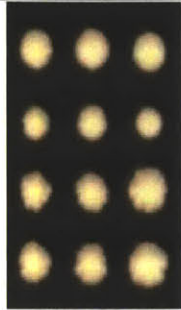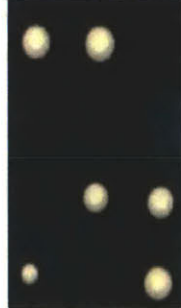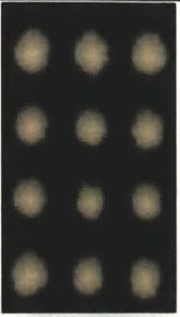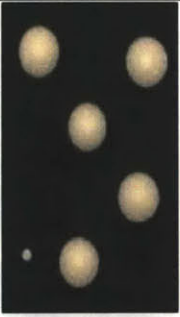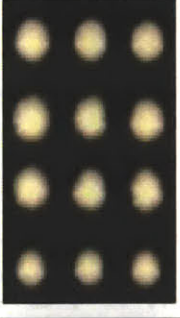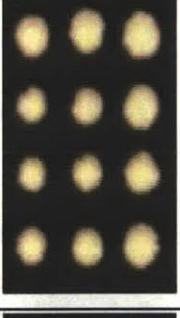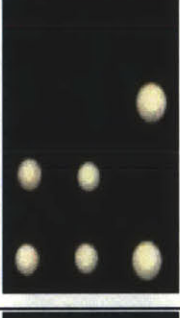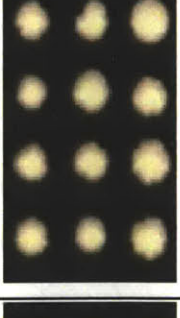
Figure S2-2 | **Chromsomal SNP comparison between S288c and Σ1278b.** Graphs comparing S288C (x-axis) chromosomes to Σ1278b. The Y-axis indicates the number of SNPs per window, with rolling windows of length 500, on alignments generated by fsa (v1.07 (BRADLEY *et al.* 2009)). Regions of large structural differences (insertions, deletions, and translocations) are indicated by dark grey boxes below the zero axis.
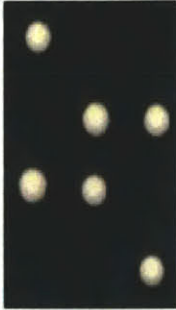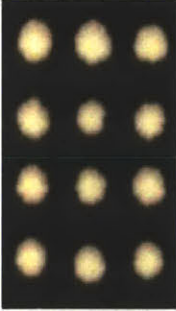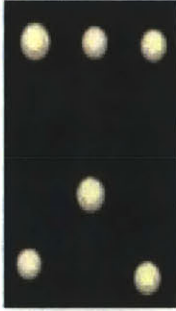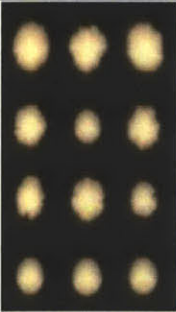
Table S2-1 | **Tetrad confirmation of strain specific essentials.** Conditional essentials were defined by tetrads in which both deletion bearing spores failed to germinate after 4 days in one strain, but both germinated when made in the other background. The suppression of the lethal ranged from excellent (growth indistinguishable on YPD from wild type) to partial.
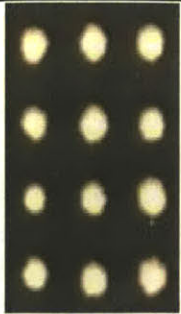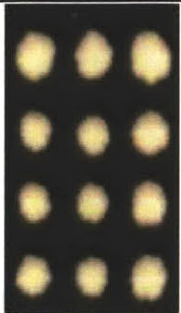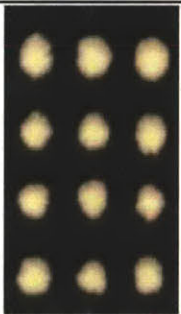
Σ1278b Specific Essentials:
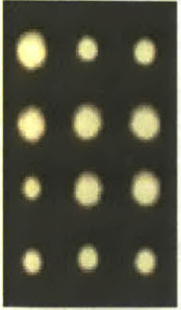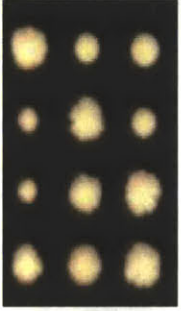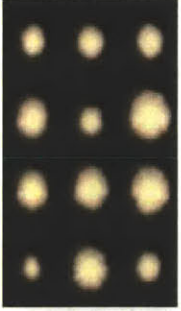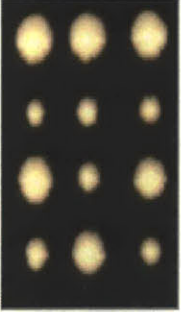
Gene	S288c Tetrads	Σ1278b Tetrads
aat2Δ		
bem1Δ		
fmp27Δ		
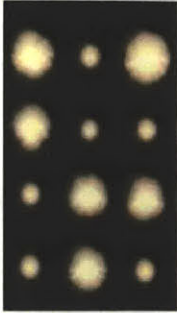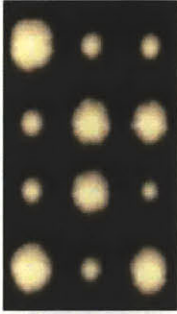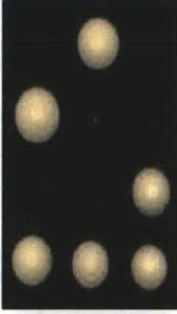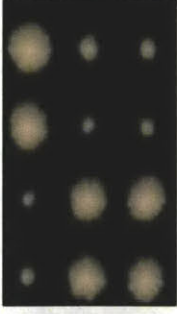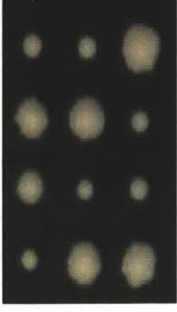
Gene	S288c Tetrads	Σ1278b Tetrads
lea1Δ		
mcm22 Δ		
mto1 Δ		
pep12Δ		
pep7Δ		
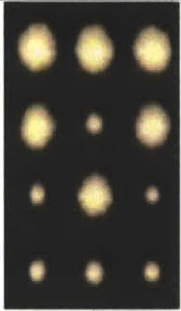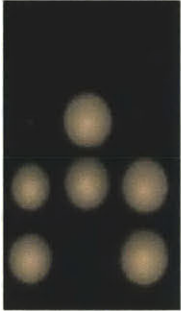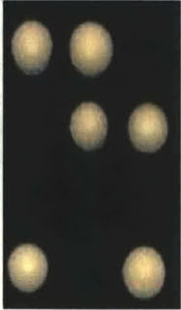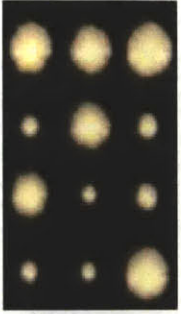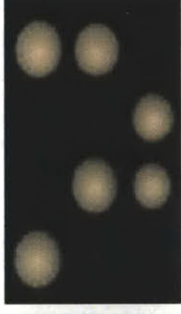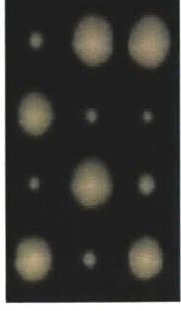
Gene	S288c Tetrads	Σ1278b Tetrads
sbh2Δ		
ski7Δ		
ski8Δ		
vps16Δ		
ydl089wΔ		
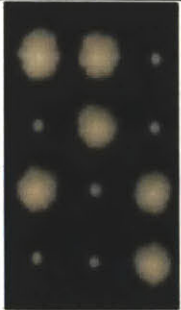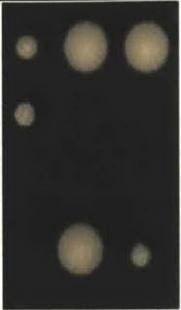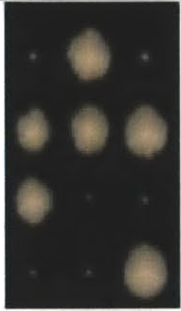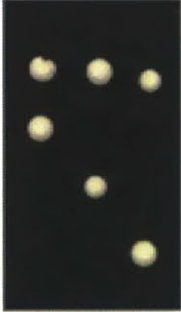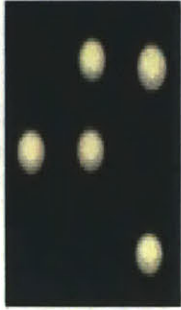
Gene	S288c Tetrads	Σ1278b Tetrads
yhr009cΔ		
ykr075cΔ		
ypr015cΔ		
zwf1Δ		
cys3Δ		

Gene	S288c Tetrads	Σ1278b Tetrads
cys4Δ		
rps10aΔ		
npl3Δ		
lsm6Δ		
pho88Δ		

Gene	S288c Tetrads	Σ1278b Tetrads
pho90Δ		
adk1Δ		
arp5Δ		
ies6Δ		
ost4Δ		

Gene	S288c Tetrads	Σ1278b Tetrads
snt309Δ		
ydr241wΔ		
lsm7Δ		
swi6Δ		
tma108Δ		
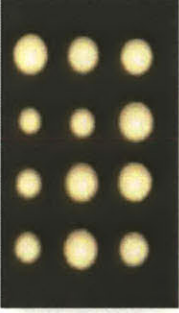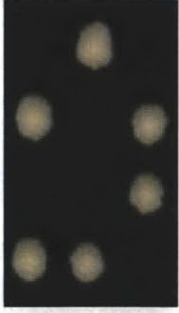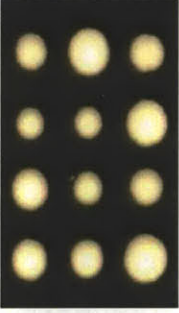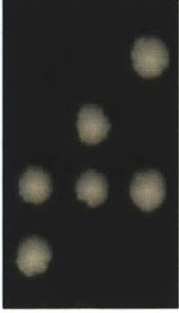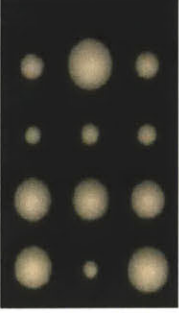
Gene	S288c Tetrads	Σ1278b Tetrads
vps34Δ		
vps75Δ		
uaf30Δ		
pop2Δ		
ctk1Δ		

Gene	S288c Tetrads	Σ1278b Tetrads
cyc8Δ		
gon7Δ		
cdc40Δ		
cgr1Δ		
rom2Δ		
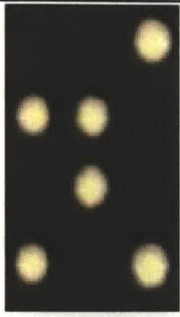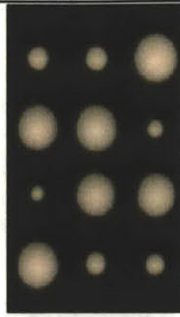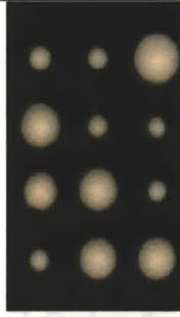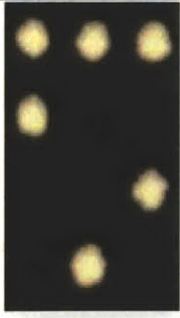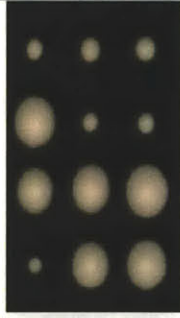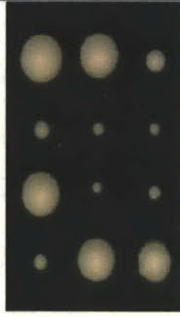
Gene	S288c Tetrads	Σ1278b Tetrads
*utr1*Δ		

S288c Specific Essentials:

Gene	S288c Tetrads	Σ1278b Tetrads
*plp2*Δ		
*ret2*Δ		
*pfy1*Δ		

Gene	S288c Tetrads	Σ1278b Tetrads
rho3Δ		
srp101Δ		
srp102Δ		
srp21Δ		
srp14Δ		

Gene	S288c Tetrads	Σ1278b Tetrads
srp72Δ		
uso1Δ		
yml6Δ		
srp68Δ		
ubc1Δ		

Table S2-2 | Strains utilized

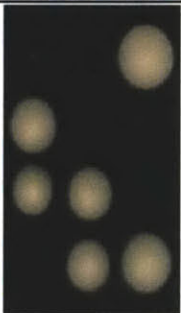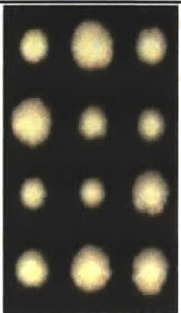| Strain	Genotype	Background	Reference
10560-6B	*MATα ura3-52 trp1::hisG leu2::hisG his3::hisG*	Σ1278b	Fink lab strain collection
YSWT3	*MATa/α can1Δ::STE2pr-sphis5/CAN1 lyp1Δ::STE3pr-LEU2/LYP1 his3::hisG his3::hisG leu2Δ/leu2Δ ura3Δ/ura3Δ*	Σ1278b	this study
Y1239	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*	S288c	Rosetta strain BY4741a
Y3295	*MATα ura3Δ leu2Δ his3::hisG*	Σ1278b	Microbia strain MT1562
Y12868	*MATa/MATα his3Δ1/his3::hisG leu2Δ0/leu2Δ0 ura3Δ0/ura3Δ0 met15Δ0/MET15*	S288c x Σ1278b	this study

Table S2-3 | **Hybrid tetrad analysis of 18 Σ1278b specific essentials** Hybrid mutant strains were created by crossing the MATa deletion mutants from the S288c collection to Σ1278b wild type (Y3295). Tetrads were dissected and scored for segregation patterns (parental ditype 2:2; nonparental ditype 4:0; and tetratypes 3:1). A chi-squared p-value was then determined to test three hypothesis: (1) a single unlinked modifier explains the in- heritance patterns (single gene p-value) when a 1:1:4 ratio is anticipated; (2) three unlinked modifiers explain the inheritance patterns (three gene p- value) when a 1:163:53 ratio is anticipated; (3) that complex genetics (multiple loci) make the inheritance patterns indistinguishable from the empirically observed background (wild type p-value). All 18 cases reject the null hypothesis (p-value < 0.01) of a single gene modifier. The null hypothesis of three modifiers is rejected by three genes (LEA1, FMP27, and YPR015C). The null hypothesis of inheritance indistinguishable from background is rejected in 5 cases. Finally, the observed wild-type frequencies also reject the single and three gene hypothesis.

Gene	Total Tetrads	Parental ditype (2:2)	Nonparental ditype (4:0)	Tetratype (3:1)	single gene p-value	three gene p-value	wild type p-value
bem1Δ	116	32	22	62	2E-3	0	1E-191
ski7Δ	89	26	21	42	3E-4	0	8E-133
lea1Δ	62	1	50	11	1E-40	0.22	1E-4
fmp27Δ	59	1	49	9	2E-41	0.12	3E-3
ypr015cΔ	49	1	41	7	3E-35	0.08	0.02
sbh2Δ	61	5	47	9	1E-35	3E-18	3E-5
pep7Δ	64	4	53	7	1E-44	7E-12	0.01
tma108Δ	61	3	50	8	2E-41	4E-7	0.01
zwf1Δ	66	3	55	8	1E-46	7E-7	0.02
mto1Δ	89	5	77	7	9E-69	3E-14	0.07
ski8Δ	61	2	53	6	9E-48	3E-4	0.27
ykr075cΔ	59	0	57	2	1E-59	8E-4	0.38
yhr009cΔ	61	2	56	3	1E-54	2E-5	0.27
aat2Δ	58	2	51	5	8E-47	1E-4	0.47
ydl089wΔ	60	2	53	5	6E-49	1E-4	0.52
pep12Δ	47	1	45	1	1E-46	6E-4	0.60
vps16Δ	63	1	60	2	1E-61	3E-4	0.67
mcm22Δ	72	1	67	4	2E-66	7E-4	0.87
wild type	129	3	119	7	5E-116	6E-8	-

All data files can be downloaded from

Data File S1: Annotation of open reading frames and noncoding RNAs in Σ1278b

The file contains the Σ1278b annotation in tab-delimited format with 9 columns: orfname, gene name, chromosome, strand, start, end, number of exons, exon starts (separated by commas), exon ends (separated by commas). The orfname utilizes the S288c ortholog when available and otherwise a Σ1278b specific systematic name.

Data File S2: Heterozygous deletion collection for Σ1278b

The file contains the Σ1278b deletion collection in tab-delimited format with 7 columns: index, orfname, gene name, set identifier, row and column location, uptag sequence, and downtag sequence. Tag sequences are given a 5 to 3.

Supplemental References

BLANCHETTE, M., W. J. KENT, C. RIEMER, L. ELNITSKI, A. F. SMIT *et al.*, 2004 Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res **14**: 708-715.

BRADLEY, R. K., A. ROBERTS, M. SMOOT, S. JUVEKAR, J. DO *et al.*, 2009 Fast statistical alignment. PLoS Comput Biol **5**: e1000392.

BROAD, Saccharomyces cerevisiae RM11-1a Sequencing Project. Broad Institute of Harvard and MIT (http://www.broad.mit.edu), pp. Broad Institute.

COSTANZO, M., A. BARYSHNIKOVA, J. BELLAY, Y. KIM, E. D. SPEAR *et al.*, 2010 The genetic landscape of a cell. Science **327**: 425-431.

DELCHER, A. L., A. PHILLIPPY, J. CARLTON and S. L. SALZBERG, 2002 Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res **30**: 2478-2483.

DICKSON, S. P., K. WANG, I. KRANTZ, H. HAKONARSON and D. B. GOLDSTEIN, 2010 Rare variants create synthetic genome-wide associations. PLoS Biol **8**: e1000294.

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

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

KELLIS, M., N. PATTERSON, M. ENDRIZZI, B. BIRREN and E. S. LANDER, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423**: 241-254.

KENT, W. J., 2002 BLAT--the BLAST-like alignment tool. Genome Res **12**: 656-664.

KORBEL, J. O., A. E. URBAN, J. P. AFFOURTIT, B. GODWIN, F. GRUBERT *et al.*, 2007 Paired-end mapping reveals extensive structural variation in the human genome. Science **318**: 420-426.

LOWE, T. M., and S. R. EDDY, 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25**: 955-964.

MAJOROS, W. H., M. PERTEA and S. L. SALZBERG, 2004 TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics **20**: 2878-2879.

PHILLIPS, P. C., 1998 The language of gene interaction. Genetics **149**: 1167-1171.

ROZEN, S., and H. SKALETSKY, 2000 Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132**: 365-386.

SALZBERG, S. L., M. PERTEA, A. L. DELCHER, M. J. GARDNER and H. TETTELIN, 1999 Interpolated Markov models for eukaryotic gene finding. Genomics **59**: 24-31.

SHERMAN, F., G. R. FINK and J. B. HICKS, 1986 Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

THOMPSON, J. D., D. G. HIGGINS and T. J. GIBSON, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22**: 4673-4680.

TONG, A. H., G. LESAGE, G. D. BADER, H. DING, H. XU *et al.*, 2004 Global mapping of the yeast genetic interaction network. Science **303**: 808-813.

WANG, J., W. WANG, R. LI, Y. LI, G. TIAN *et al.*, 2008 The diploid genome sequence of an Asian individual. Nature **456**: 60-65.

ZERBINO, D. R., and E. BIRNEY, 2008 Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:** 821-829.

Chapter 3

Genetic variation within a species: Circuit diversification in a signal transduction network

Brian L. Chin[1], Owen Ryan[2,3,4], Fran Lewitter[1], Charles Boone[2,3,4], Gerald R. Fink[1,5]

[1]Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

[2]Banting and Best Department of Medical Research, University of Toronto, Toronto, ON M5S 3E1, Canada.

[3]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 3E1, Canada.

[4]Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada.

[5]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA

Abstract

Construction of identical deletion libraries in two closely related strains of yeast permits the assessment of phenotypic differences between individuals of a single species for the same mutation. Here, we compare the reference strain S288c with a related strain, Sigma, for the ensemble of genes that affect a morphological trait, adhesion/filamentation. The nucleotide divergence between the two strains is roughly equal to that between the genomes of two humans in the population. Previous studies, all in Sigma, had shown that this trait was controlled by the filamentation MAPK (fMAPK) pathway, which activates a set of transcription factors required for the transcriptional activation of a downstream structural gene *FLO11*. Unexpectedly, the fMAPK pathway is not required to activate *FLO11* for adhesion/filamentation in S288c despite the fact that its kinases are present and active in other pathways. In S288c the requirement for the fMAPK pathway is partially bypassed by a polymorphic transcription factor *RPI1*. The *RPI1* allele from S288c but not the one from Sigma can confer MAPK pathway independence. The two alleles differ in the number of tandem repeats in the ORF. Repeat length polymorphisms in numerous orthologous ORFs are frequent in closely related strains and result in enormously high levels of variation in progeny genotypes. Thus, even within a species there can be substantial differences in the networks that control gene expression.

Introduction

Even though recent advances in DNA sequencing have identified many nucleotide polymorphisms, association of genetic variation with specific phenotypic differences among individuals has been difficult for complex traits (DICKSON *et al.* 2010; JAKOBSDOTTIR *et al.*

2009; MANOLIO *et al.* 2009). This difficulty has been variously attributed to both genetic and non-genetic factors (CARLBORG and HALEY 2004; DICKSON *et al.* 2010; HARTMAN *et al.* 2001; KORBEL *et al.* 2007). Among the genetic origins are: many genes contributing a small effect to the final phenotype and complex (epistatic) gene interactions. Due to the complexity of the problem, the baker's yeast *S. cerevisiae*, with its compact and tractable genome, is a good choice to work out the principles underlying the genotype to phenotype problem.

Early sequence studies in yeast focused on comparing *S. cerevisiae* to other yeast species diverged by as much as 20 million years (KELLIS *et al.* 2003). While these studies have been important in aiding our understanding of yeast evolution, they have not addressed how small genetic differences impact phenotypes. Subsequent sequencing studies have examined large numbers of *S. cerevisiae* strains from a variety of sources, but have focused on population structure and evolutionary origins of the strains rather than the problem of connecting genotype to phenotype (LITI *et al.* 2009; SCHACHERER *et al.* 2009).

While large sequencing studies have generally not addressed the impacts of natural variation in the *S. cerevisiae* population, linkage studies dealing with a few strains have yielded important insights about the amount of variation within the species. The cross of the wild vineyard strain RM11 to the standard laboratory strain S288c has been particularly influential in showing how variants in the genome can lead to phenotypic differences. Traits examined in this cross include gene expression, morphology, resistance to DNA damaging agents, and telomere length (BREM *et al.* 2002; DEMOGINES *et al.* 2008; GATBONTON *et al.* 2006; NOGAMI *et al.* 2007). These studies have identified some loci relevant to the phenotype, but in all cases the genetic basis for the observed differences between strains was complex. Even with sufficient statistical

power to identify the many polymorphisms influencing the phenotype, the mechanisms responsible for the divergent phenotypes were not apparent.

To address these issues, recent studies developed a model system that enables a comprehensive assessment of phenotypic differences for the same mutation in the two genetic backgrounds S288c and Σ1278b (Sigma) (DOWELL *et al.* 2010). The two strains have very similar genomic sequences: their divergence of ~0.3% is similar to that between unrelated humans. To assess functional differences between these two strains, ~5100 genes were deleted in Sigma for comparison with the same set of deletions in S288c (DOWELL *et al.* 2010; WINZELER *et al.* 1999). The initial genome-wide survey identified conditionally essential genes, genes whose function is required for viability in one strain but not the other. The analysis indicated that the basis for the conditional essentiality was likely a complex set of background modifiers, and in many cases 5-6 modifiers appeared to be present in one strain and not the other.

Here we have also shown that these two strains differ dramatically in the circuitry required for a key morphogenetic trait, adhesion/filamentation. A comparison of the two genomic deletion libraries showed that the filamentation mitogen activated kinase (fMAPK) pathway required for the transition from yeast form to filamentous form in the Sigma strain is not required in S288c. Instead, S288c has mechanisms to bypass the fMAPK pathway and activate the transcription of *FLO11*. Among the many modifiers, we identified the transcription factor *RPI1* as one bypass suppressor. Only the *RPI1* allele from S288c (*RPI1*^S288c^) confers fMAPK pathway independence, and it can do so in either genetic background. The Sigma allele (*RPI1*^Sigma^) fails to activate *FLO11* transcription in either strain. Moreover, the *RPI1*^S288c^, but not the *RPI1*^Sigma^ is hyperphosphorylated both in S288c and Sigma. The two forms of *RPI1* differ in the number of tandem repeats in the ORF. A genome-wide comparison of these strains shows

that many other genes with intragenic tandem repeats are highly polymorphic with respect to repeat size. Thus, changes in internal repeat number, which arise frequently (VERSTREPEN *et al.* 2005), can lead to major changes in gene expression and provide a wealth of opportunities for evolutionary selection.

Results

The fMAPK pathway is required for adhesion and *FLO11* transcription in Sigma but not in S288c^{FLO8}

Systematic comparison between S288c and Sigma for the adhesion phenotype required the creation of a new S288c^{FLO8} library because the progenitor to the standard S288c deletion library carries a *flo8* mutation that prevents adhesion to agar and other surfaces. When S288c is made wild-type for *FLO8* (S288c^{FLO8}), it adheres in a *FLO11* dependent fashion(LIU *et al.* 1996).

Screening the S288c^{FLO8} library and the comparable Sigma deletion library for adhesion uncovered 599 deletions with decreased adhesion (Ahs⁻) (Supplemental Tables 3-1 and 3-3). Only 46 deletions affected adhesion the same way in both strains but the fMAP kinase pathway was not present in this set of shared adhesion functions. Instead, the fMAPK pathway is required for adhesion only in Sigma and is not required for adhesion in S288c^{FLO8}. Strains carrying deletions in kinase genes, *STE7*, *STE11*, *KSS1*, and the transcription factor genes, *STE12*, and *TEC1*, have a clear adhesion defect in Sigma but adhere well in S288c^{FLO8} (Figure 3-1A). qPCR measurements revealed that the wild-type S288c^{FLO8} and S288c^{FLO8} *tec1*Δ both show strong expression of *FLO11*, whereas Sigma *tec1*Δ has a 50-fold decrease in *FLO11* RNA levels relative to the wild-type control (Figure 3-1C). The distinct requirement for the fMAPK

71

pathway in Sigma but not in S288c^{FLO8} suggests that adhesion is controlled differently in the two strains.

The fMAPk pathway in Sigma activates *FLO11* transcription for haploid adhesion and diploid filamentation (LIU *et al.* 1993; LO and DRANGINIS 1998; ROBERTS and FINK 1994). To determine whether the fMAPK pathway is dispensible for diploid filamentation in S288c^{FLO8}, we constructed diploid S288c^{FLO8} strains. Filamentation in the S288c^{FLO8} *tec1Δ/tec1Δ* strain is indistinguishable from wild-type, whereas the Sigma *tec1Δ/tec1Δ* strain has a filamentation defect (Figure 3-1B). A hybrid S288c^{FLO8}/Sigma *tec1Δ/tec1Δ* strain is able to filament, showing that the ability of S288c^{FLO8} to bypass an fMAPK defect for filamentation is dominant. Homozygous diploid S288c^{FLO8} *flo11Δ/flo11Δ* or Sigma *flo11Δ/flo11Δ* strains failed to form filaments. Thus, *FLO11* function is required for adherence and filamentation in both S288c^{FLO8} and Sigma even though the requirement for the fMAPK pathway is restricted to Sigma.

fMAPK independent expression of S288c^{FLO8} *FLO11* is not due solely to strain-specific differences in the *FLO11* promoter

Reciprocal promoter swap strains were used to determine whether the sequence differences between the S288c and Sigma *FLO11* promoters (*FLO11pr^{S288c}* and *FLO11pr^{Sigma}* respectively), could account for the fMAPK independence of S288c^{FLO8}. S288c^{FLO8} *FLO11pr^{Sigma}* adhered like a wild-type S288c^{FLO8} as did S288c^{FLO8} *FLO11pr^{Sigma} tec1Δ* This shows that *FLO11pr^{S288c}* is not necessary for fMAPK independent adhesion of S288c cells (Figure 3-2). *FLO11* RNA levels in the S288c^{FLO8} *FLO11pr^{Sigma}* strain were consistent with the adhesion phenotypes. Specifically, in S288c^{FLO8} there was no significant difference in *FLO11* RNA levels, regardless of the promoter, or the presence of a *tec1Δ* (Supplemental Figure 3-1A).

The $FLO11pr^{S288c}$ does not promote $FLO11$ transcription as efficiently in Sigma as it does in S288c. This difference is reflected both in the adhesion assay and the qPCR measurement of $FLO11$ RNA levels. Nevertheless, the $FLO11pr^{S288c}$ in Sigma is $TEC1$ dependent for both adhesion and $FLO11$ transcription, whereas it is $TEC1$ independent in $S288c^{FLO8}$ (Figure 3-2 and Supplemental Figure 3-1B). These results imply that the sequence differences in the promoters are not responsible for the fMAPK independence of $S288c^{FLO8}$.

The strain difference in $FLO11$ regulation is genetically complex

Since the $FLO11pr^{S288c}$ is neither necessary, nor sufficient for fMAPK independent adhesion, we sought to determine how many allelic differences between S288c and Sigma affect the difference in adhesion phenotype between these two strains. We crossed the adherent $S288c^{FLO8}$ $tec1\Delta$ strain to the non-adherent Sigma $tec1\Delta$, and examined the adhesion properties of 24 complete meiotic tetrads (Supplemental Figure 3-2). 22/96 progeny were clearly adherent, 56/96 were non-adherent, and 16/96 displayed an intermediate phenotype. Although the intermediate phenotypes in the F1 progeny make it difficult to score the adhesion phenotype with certainty, the ratios of adhesive to non-adhesive progeny suggest that ≥ 3 genes play a role in fMAPK independent adhesion. Backcrosses of the adherent progeny to the non-adherent Sigma $tec1\Delta$ parent failed to elucidate the number of genes controlling the adherence difference: Despite three generations of directed backcrossing, the segregation patterns remained complex and precluded identification of specific loci that permit fMAPK independent adhesion.

The complex inheritance is not a consequence of the method for scoring adherence/non-adherence. When a chromosomal $FLO11pr::GFP$ construct was used to monitor the segregation

of *FLO11* expression in S288c Δ*tec1* x Sigma Δ*tec1* crosses, the same complex inheritance pattern reflecting *FLO11* transcription was observed (Supplemental Figure 3-3).

Dominant S288c modifiers confer fMAPK independent expression of *FLO11*

To identify the S288c polymorphisms responsible for fMAPK independent adhesion, we developed a Sigma transformation protocol to select for plasmids carrying S288c genes that bypass the fMAPK pathway. The selection utilized a Sigma *tec1*Δ strain in which the *FLO11* ORF was replaced with a *HIS3-PEST* construct. Because Sigma requires Tec1p to regulate *FLO11*, the Sigma *FLO11pr-HIS3-PEST*, *tec1*Δ strain is His⁻ whereas the S288c^{FLO8} *FLO11pr-HIS3-PEST*, *tec1*Δ strain is His^{+}. Genes from S288c that could bypass the requirement for the fMAPK pathway in Sigma were obtained by transforming the Sigma *FLO11pr-HIS3-PEST*, *tec1*Δ strain (His⁻) with a S288c CEN/ARS genomic library (ROSE *et al.* 1987) and selecting for His^{+} transformants.

Four genes were identified as bypass suppressors of *tec1*Δ: *TEC1*, *RPI1*, *MIT1*, and *MGA1*. As the *MGA1* gene lacks sequence or expression differences between S288c and Sigma and suppresses *tec1*Δ when expressed from a high copy number plasmid, it was not pursued further (LORENZ and HEITMAN 1998). However, comparison of the S288c DNA sequences of *RPI1* and *MIT1* with their cognates in Sigma revealed many differences. Both *RPI1* and *MIT1* contain numerous SNPs and stretches of intragenic repeats that differ in length between S288c and Sigma (Figure 3 - 3 and Supplemental Figure 3-4). Because of these differences we further examined the role that *RPI1* and *MIT1* play in *FLO11* regulation.

RPI1^{S288c} but not the *RPI1*^{Sigma} is a bypass suppressor of the fMAPK pathway

Consistent with the hypothesis that the $RPI1^{S288c}$ has an allele specific role in $FLO11$

expression, deletion of $RPI1^{S288c}$ in S288c^{FLO8} results in a strong adhesion defect and decreased

$FLO11$ RNA, whereas deletion of $RPI1^{Sigma}$ in Sigma does not (Figure 3-4A and 3-4B). By

contrast, deletions of $MIT1^{S288c}$ in S288c^{FLO8} or of $MIT1^{Sigma}$ in Sigma both manifest a strong

adhesion defect and decreased $FLO11$ RNA. These data suggest that despite the sequence

differences, $MIT1$ has no allele specificity with respect to the phenotypes we have assayed and

may be a high copy suppressor.

To further characterize $RPI1$ allele specificity, we constructed an S288c^{FLO8} strain where

the $RPI1$ sequence was replaced by $RPI1^{Sigma}$, and reciprocally, a Sigma strain where the

$RPI1^{Sigma}$ sequence was replaced by $RPI1^{S288c}$. S288c^{FLO8} $RPI1^{Sigma}$ displayed an adherence

phenotype and $FLO11$ RNA levels that were not significantly different than an $rpi1\Delta$, suggesting

that $RPI1^{Sigma}$ is not functional in $FLO11$ regulation (Figure 3-4A and 3-4B). Deletion of $TEC1$

in S288c^{FLO8} $RPI1^{Sigma}$ does not further decrease adhesion or $FLO11$ levels. Reciprocally, the

Sigma $RPI1^{S288c}$ strain had $FLO11$ mRNA levels that were comparable to wild-type, and Sigma

$RPI1^{S288c}$ $tec1\Delta$ showed more $FLO11$ RNA than the Sigma $RPI1^{Sigma}$ $tec1\Delta$, but less than wild-

type (Figure 3 - 4C). These results show that the $RPI1^{S288c}$ allele promotes $FLO11$ expression

and can partially bypass the $tec1\Delta$; however, the $RPI1^{Sigma}$ allele is unable to bypass $tec1\Delta$.

Rpi1p interaction with the $FLO11$ promoter is Rpi1p allele specific

To determine whether the difference in fMAPK independent $FLO11$ expression is due to

differences in the ability of Rpi1p^{Sigma} and Rpi1p^{S288c} to interact with the $FLO11$ promoter, we

performed chromatin immunoprecipitation (ChIP) by using 3xFLAG tagged alleles of Rpi1p,

and tested for enrichment of the $FLO11$ promoter. Rpi1p^{S288c} interacts with the $FLO11$

promoter with a peak around -1300bp (Figure 3-5A), the site where other positive activators of

FLO11 such as Tec1p, and Flo8p bind (Borneman *et al.* 2006; Zeitlinger *et al.* 2003).

Immunoprecipitation of the Rpi1p^{S288c} allele enriches for the *FLO11* promoter regardless of the

strain background. In contrast to Rpi1p^{S288c}, immunoprecipitation of the Rpi1p^{Sigma} results in

strain-background-specific enrichment for this same region of the *FLO11* promoter. When the

Rpi1p^{Sigma} is immunoprecipitated from a Sigma strain, it enriches for the *FLO11* promoter; when

it is immunoprecipitated from an S288c^{FLO8} strain it does not.

This difference between Rpi1p^{S288c} and Rpi1p^{Sigma} promoter binding is also observed at

the promoter of *MIT1*, previously identified as a target of Rpi1p and a "master regulator" of

FLO11 transcription (Cain *et al.* 2011; Wang *et al.* 2011). However, Wang et al. and Caine et

al. provided only a strain specific analysis of *MIT1* and *RPI1* function: The Mit1p^{Sigma} protein

was shown to bind to the *FLO11* promoter in Sigma, and Rpi1p^{S288c} has been reported to localize

to the promoter of *MIT1*^{S288c} in S288c. Our ChIP data show that Rpi1p^{S288c} localizes to the *MIT1*

promoter, regardless of strain background, but Rpi1p^{Sigma} localizes to the *MIT1* promoter only in

the Sigma background (Figure 3-5B). Furthermore, Rpi1p^{S288c} requires a functional *MIT1* to

suppress a defect in the fMAPK pathway in both S288c and Sigma. Rpi1p^{Sigma} can interact with

both the *FLO11* and *MIT1* promoters in Sigma, but not in S288c^{FLO8}. Thus, Rpi1p^{Sigma} must be

structurally different from Rpi1p^{S288c} and require additional factors to function.

The Rpi1 protein is differentially phosphorylated in the two strains

Analysis of the Rpi1p protein showed that Rpi1p^{S288c} is structurally different from

Rpi1p^{Sigma}. Figure 7 shows that 3x-FLAG-tagged Rpi1p^{S288c} extracted from S288c and

visualized on Western blots runs as a diffuse species different from the Rpi1p^{Sigma} band from

Sigma. When the Rpi1p^{S288c} is expressed in Sigma, it again runs as a diffuse higher molecular weight species, but when Rpi1p^{Sigma} is expressed in S288c^{FLO8}, it runs as a single band (Figure 3-6).

To determine if the difference between the isoforms of Rpi1p is due to phophorylation, protein extracts were treated with lambda phosphatase. The Rpi1p^{S288c} smear collapsed to a single band. This change in migration pattern occurs regardless of the strain background that expresses Rpi1p^{S288c}. Treatment of Rpi1p^{Sigma} with phosphatase changed its migration on the gel only if the protein came from the Sigma strain. These experiments show that Rpi1p^{Sigma} has strain-specific phosphorylation and likely has a different phosphorylation pattern than Rpi1p^{S288c}. This altered phosphorylation pattern of Rpi1p^{Sigma} may account for its inability to activate *FLO11* transcription in either strain (Figure 3-4B and 3-4C).

Discussion

Individuals within a species may signal gene expression through different pathways

Analyses of related species have shown that gene regulatory networks can be rewired. Such changes have been documented in the control of mating type regulation between the fungi *S. cerevisiae* and *Candida albicans*. Both species have MATa cells that express a-specific genes (asgs) and Matα cells that express α-specific genes (αsgs). While both species use orthologous regulators to perform this task, the regulatory mechanisms differ significantly. *C. albicans* has the ancestral mode of regulation where expression of asgs are off by default and their expression must be activated by the HMG domain protein a2 (CAIN *et al.* 2011; TSONG *et al.* 2003; TSONG *et al.* 2006). This process has been rewired in *S. cerevisiae* which has lost a2. In *S. cerevisiae* the expression of the asgs is on by default and they must be repressed by the homeodomain protein α2. This switch from positive regulation to negative regulation required both changes in cis and trans regulatory elements, but how this rewiring can occur without a loss of fitness is difficult to imagine.

The divergence of regulatory pathways between species such as *S. cerevisiae* and *C. albicans* allows a look into the outcomes of evolution, but it is difficult to know the processes that led to the divergence of the pathways. Our analysis of comparable deletion libraries in two inter-fertile strains of *S. cerevisiae* found that regulatory differences can exist within a species due to the natural variation between individuals.

Our finding that that Sigma requires the fMAPK pathway for *FLO11* expression, whereas S288c^{FLO8} does not, was unanticipated because the fMAPK pathway is conserved among evolutionarily distant fungal species (LANE *et al.* 2001). The S288c^{FLO8} bypass of the fMAPK

pathway is partly explained by the existence of an allele of the polymorphic transcriptional activator $RPI1^{S288c}$.

While $RPI1^{S288c}$ can bypass the fMAPK pathway, $FLO11$ expression and genetic analysis suggests that the $RPI1^{S288c}$ allele is not the only gene able to bypass the fMAPK pathway for $FLO11$ expression. When integrated in Sigma, $RPI1^{S288c}$ only partially restores the fMAPK defect for $FLO11$ expression. S288c^{FLO8} can also bypass the fMAPK pathway on either YPD or synthetic media, but the contribution of $RPI1^{S288c}$ to Sigma adhesion is most pronounced when Sigma is grown on synthetic media, the condition under which it was selected. Even under these conditions, the bypass seen in a Sigma $RPI1^{S288c}$ $tec1\Delta$ is not as strong as in S288c^{FLO8} $tec1\Delta$. These differences suggest that, while $RPI1^{S288c}$ clearly bypasses the fMAPK pathway, it is only one of several S288c polymorphisms that bypass the fMAPK pathway. Over evolutionary time, the presence of the $RPI1^{S288c}$ allele in conjunction with these other modifiers has the potential to free the fMAPK pathway to control other targets without losing the ability to undergo the key morphogenetic switch to adhesion/filamentation.

Evolution of divergence within a species

These results emphasize that considerable variation exists in the circuitry of key signaling pathways even among members of the same species. All isolates of the standard reference strain S288c have a nonsense mutation in the $FLO8$ gene, and are unable to adhere or filament. But, S288c $flo8$ isolates are heterogeneous: some have a non-functional $KSS1$ gene encoding the filamentation specific MAPK (ELION et al. 1991). The existence of both S288c $flo8$, $kss1$ and S288c $flo8$, $KSS1$ strains supports the idea that the fMAPK-specific members are not necessary in S288c. The elements of the fMAPK pathway that have been conserved (Ste20p, Ste11p,

Ste7p, Ste12p) are under strong positive selection in the laboratory because they function in additional signal transduction pathways (mating, osmotic-sensing). In fact, circuitry of the mating and hyper-osmolarity glycerol pathways was uncovered through genetic analysis of S288c strains.

The finding that $RPI1^{S288c}$ is active in $FLO11$ regulation suggests that it functions in conjunction with many pathways. For example, $FLO11$ transcription in Sigma requires a number of previously identified transcriptional regulators, including Mga1p, Phd1p, Sok2p, Ste12p, Tec1p, and Flo8p, and genome-wide ChIP analyses have shown that these six regulators regulate each other and a network of hundreds of downstream targets (BORNEMAN et al. 2007a; BORNEMAN et al. 2006; BORNEMAN et al. 2007b; MONTEIRO et al. 2008). The finding that $RPI1$ binds to the $MIT1$ promote r(WANG et al. 2011), $MIT1$ itself being a transcriptional activator of many genes (CAIN et al. 2011) is consistent with the complex network that regulates $FLO11$. $RPI1$ is a newly discovered transcriptional activator of $FLO11$ and its polymorphisms add a new dimension to the complexity of the regulatory network controlling $FLO11$ expression.

The discovery of $RPI1^{S288c}$ as a bypass suppressor provides an example of the mechanism by which allelic polymorphisms can buffer the effect of mutations. The presence of $RPI1^{S288c}$ in S288c means that loss of function of any member of the fMAPK pathway will fail to manifest an adhesion phenotype. In a therapeutic context this variation could explain why a drug directed towards some elements of a conserved signaling pathway that has gone awry in the diseased state may be ineffective in some individuals.

Intragenic tandem repeats are highly polymorphic within a species

Although the two *RPI1* alleles differ by several nucleotide changes, the most striking is the alteration in the size of a repeat region present in the coding sequence of the gene (Figure 3 - 4). Repeat polymorphisms in *RPI1* are present in wild isolates of yeast as well as in many laboratory strains, and previous studies have suggested that changes in the length of internal tandem repeats can have phenotypic consequences (FIDALGO *et al.* 2008; LEVDANSKY *et al.* 2007; MACDONALD *et al.* 1993; SHEETS and ST GEME 2011; TAN *et al.* 2010; VERSTREPEN *et al.* 2005).

Repeats within a coding sequence create enormous flexibility for the evolution of diversity within a species. Base pair mutations occur at a frequency of 10^{-6}, are usually deleterious, and revert to a functional protein only at 10^{-8}. By contrast expansion/contraction of tandem repeats can occur ~1000x more frequently. For this reason, they are highly mutable, capable of creating novel functions at high frequency. These protean elements provide the plasticity that enables a species to adapt to many environmental conditions without becoming irreversibly committed to a phenotype.

Moreover, changes in intragenic repeats create unanticipated genome diversity even within a single species. The sequence divergence between *S. cerevisiae* and its closest relative, *S. paradoxus* is 20%. However, these comparisons do not take into account the diversity of intragenic repeats and most *Saccharomyces* genome assemblies do not accurately determine the number of elements within repeat regions. Based on the finding of repeat length changes in *RPI1* and *MIT1*, a genome-wide comparison of ORFs between S288c and Sigma revealed 107 genes differ in size between S288c and Sigma because they contain in-frame expansions or contractions of intragenic repeat sequences (Supplemental table 3-4). The set of genes with intragenic repeat length differences includes genes involved in diverse biological processes,

including transcription (including the general transcription factor TFA1, the TFIIE large subunit, involved in recruitment of RNA polymerase II to the promoter), chromatin modification, and signal transduction. To ensure that these differences are not due to sequencing errors, 24 of these length differences were verified by PCR (Figure 3-8 and Supplemental Figure 3-5). Twenty-two of 24 genes show the predicted size difference (Figure 3-8 and Supplemental Figure 3-5), confirming the size differences predicted from the genome sequences reflecting length differences in the repeats. These data suggest that in a cross between S288c and Sigma these size polymorphisms could generate as many as 2^{100} genotypes. Realization of even a tiny fraction of this variation would provide ample grist for evolution's mill.

Acknowledgements

Material and Methods

Strains, Media, Microbiological Techniques, and Growth Conditions

Yeast strains used in this study are derived from S288c and Σ1278b. Standard yeast media were prepared and genetic manipulation techniques were carried out as described (GUTHRIE and FINK 2002).

To construct the S288c^{FLO8} deletion library, each of the 4705 deletion strains in the standard S288c *flo8* library was transformed with a CEN/ARS plasmid carrying the Sigma *FLO8* gene under the control of its own promoter using previously published methods (VOYNOV *et al.* 2006). The 4633 *FLO8* deletion strains successfully recovered from these transformations formed the S288c^{FLO8} deletion library.

Adhesion assays were carried out by densely patching strains onto YPD or SC plates. These were grown overnight at 30°C and then replica plated onto YPD or SC plates. The replica plates were grown at 30°C for three days. The S288c^{FLO8} strain expresses *FLO1* which leads to flocculation that can influence agar adhesion phenotypes. To compare agar adhesion between S288c^{FLO8} and Sigma, which does not express *FLO1*, after three days the plates were washed by partially filling petri dishes with 10mM EDTA (which disrupts *FLO1* dependent aggregates) and gentle shaking at approximately 75rpm on an orbital shaker. To visualize the difference between the strains, the media used for both the adhesion and transcription assays was optimized for intrinsic growth differences between S288c^{FLO8} and Sigma (e.g. flocculation and mother-daughter cell separation). However, the controls intrinsic to each experiment always permitted a comparison between strains grown under the same media conditions. To induce pseudohyphal growth, single cells were microdissected and grown on SLAD media which is required to visualize this morphology (GIMENO *et al.* 1992).

The strain specificity of the Ahs⁻ phenotypes is not attributable to the fact that the Sigma library had an integrated *FLO8* gene, whereas the S288c^{FLO8} library carried *FLO8* on a plasmid: The Ahs⁻ phenotype was the same in 28/30 deletions tested from the S288c deletion library whether *FLO8* was plasmid-borne or integrated at the resident *FLO8* locus (replacing the *flo8* allele). All strains used after the initial adhesion screen had the *FLO8* gene integrated at its native locus in S288c, eliminating the possibility that the phenotypic differences were a consequence of plasmid versus chromosomal expression.

For qPCR and ChIP, cells were grown overnight in liquid media as noted, diluted to OD600=0.25, and grown to OD600=4-4.5. For protein preparations, cells were grown as for qPCR in synthetic complete media.

Yeast strains carrying gene deletions were constructed by PCR amplification of kanamycin-resistance gene cassettes from the yeast deletion library (WINZELER *et al.* 2000) with approximately 200 bases of flanking sequence. Correct integrants were identified by PCR, with the exception of *tec1Δ*, which was additionally checked by Southern blot using standard techniques (BROWN 2001). *FLO11* promoter swaps were carried out by first deleting the *FLO11* promoter with the *URA3* cassette. The reciprocal swap was carried out by PCR amplifying the sequences from each strain and using the PCR products to transform the opposite strain from which the sequence was amplified. The same procedure was performed for the *RPI1* swaps but with only the ORF sequences. 3xFLAG tagged constructs were created by amplifying the *URA3* cassette from PRS306 using a primer (BCP534) that contained the 3x FLAG epitope. This construct was then subject to another round of PCR to add 50bp of flanking homology to the *RPI1* c-terminus. The resulting PCR product was used for transformation. The haploid *MATa*

deletion collection was transformed with plasmid pHL1 using previously published protocols

(LIU *et al.* 1996; VOYNOV *et al.* 2006).

GFP measurements

Cultures for GFP measurements were grown overnight in liquid YPD in 96 well plates and then pelleted and suspended in water. Samples were transferred to Corning 96 well black clear-bottom plates and OD600 and GFP fluorescence was measured in a Tecan Safire2 plate reader. For backcrosses, high fluorescing progeny were backcrossed to the low fluorescing Sigma *tec1*Δ for three generations.

*tec1*Δ bypass screen

The *CLN2* PEST sequence was added to the end of the *HIS3* gene to target the protein product to the proteasome. Without this modification, a Sigma *FLO11pr-HIS3, tec1*Δ strain produces enough His3p protein from the *FLO11* promoter to be His$^+$, even in relatively high concentrations of the His3p competitive inhibitor 3-aminotriazole. The *HIS3-PEST* construct was created by Infusion PCR cloning (Clontech) the PEST sequence from *CLN2* immediately upstream of the *HIS3* stop codon in PRS315. The *CLN2* PEST sequence was amplified using primers BCP316 and BCP317 and PRS315 was linearized by PCR using primers BCP320 and BCP321. To create the *FLO11pr-HIS3-PEST* strain, the *HIS3-PEST* construct was PCR amplified with primers BCP249 and BCP324. These primers have homology to replace the endogenous *FLO11* ORF with the *HIS3-PEST* ORF, and the PCR product was transformed into yBC172. Transformants were selected on -HIS media and then correct transformants were screened for by PCR. *TEC1* was deleted in *FLO11pr-HIS3-PEST* transformants by PCR

transformation.

The *FLO11pr-HIS3-PEST*, *tec1Δ* strains was transformed with an S288c CEN/ARS

genomic library (ROSE *et al.* 1987). Transformants were first selected for 24 hours on -URA

plates, and then replica plated onto -URA, -HIS plates plus 5mM 3-AT.

We obtained approximately 300 His$^+$ transformants out of over 15,000 total

transformants, we examined if the His$^+$ phenotype was dependent upon the plasmid by selecting

for strains that had lost the plasmid on 5-FOA. After 5-FOA selection, these strains were

examined, by dilution series, on -HIS plates.

54 strains required the library plasmid to be His$^+$, and the plasmid from these strains was

isolated and the ends of the insert were sequenced. Potential bypass strains were identified by

examining the overlapping regions among the inserts.

qPCR

Total RNA was obtained by standard acid phenol extraction from 2 ml of culture. The

Qiagen QuantiTect Reverse Transcription Kit was used to remove residual genomic DNA and

reverse transcribe the RNA templates to generate cDNAs. Aliquots of cDNA were used in Real

Time PCR analyses with reagent from Applied Biosystems and the ABI7500 real-time PCR

system.

Chromatin IP

Protocols have been described [41]. Briefly, IPs were performed with Dynal Protein G

magnetic beads pre-incubated with antibodies against FLAG-epitope (Sigma M2). To examine

enrichment, SYBR Green qPCR (Applied Biosystems) was performed on IP and WCE using gene specific primers.

Protein manipulations

Total protein was extracted using standard TCA precipitation with slight modifications [42]. Namely, after TCA precipitation the acetone wash was omitted and instead the cells were washed once with 1M Tris-pH8. For phosphatase assays, 5μl of total protein was treated with 2 μl (x units) lambda phosphatase (NEB) for 2 hours at 30°C and the reaction was stopped by adding 6x Laemmli loading buffer to 1x concentration and boiling for 10 minutes. Samples were run out on a 10% TGX gel (BioRad 456-1036S). Blotting against FLAG was performed using HRP-conjugated anti-FLAG M2 antibody (Sigma A8592).

Bioinformatics

Gene ontology term enrichment was performed using the AMIGO term enrichment tool version 1.8 (http://amigo.geneontology.org/cgi-bin/amigo/term_enrichment).

To find intragenic repeats, the EMBOSS program ETANDEM (RICE et al. 2000) was used to screen the sequences of all *S. cerevisiae* (S288c version 2010 downloaded from SGD in April 2011) and the Σ1278b strain (Sigma downloaded from http://mcdb.colorado.edu/labs1/dowelllab/pubs/DowellRyan/ in October 2010) for repeat units of length 3 to 500 bp. For each ORF, we compared the length in the two strains. We screened 6685 ORFs in S288c and 6450 ORFs in Sigma. A total of 6439 ORFs were common to both strains. Of these 6439 ORFs, 5928 were identical in length. Of the remaining 511 ORFs, 127 ORFs differed in total length by at least 6 bp and showed a length difference in the repeat region

of at least 6 bp. We eliminated an additional 11 ORFs because of large truncations in either the 5' or 3' region of the ORF accounting for the length differences between strains. All but 9 of the length differences in the 116 ORFs were a multiple of 3. These discrepancies could be due to sequencing errors. The length of the ORF was longer in Sigma for 60 ORFs (43 ORFs with bp difference of 6 to 33, 17 ORFs with bp differences of 36 or greater). A total of 56 ORFs were longer in S288c (43 ORFS with bp difference of 6 to 33, 13 ORFs with bp differences of 36 or greater).

Repeat length PCRs

Primers flanking the repeat region were designed using PRIMER3 (ROZEN and SKALETSKY 2000). PCR products were visualized on 10% polyacrylamide gels.

Figure 3-1 | **The fMAPK pathway is not required for *FLO11* expression in S288c^{FLO8}.** (A) Adhesion assays performed on Sigma strains (left half of the plate), or S288c^{FLO8} strains (right half of the plate). The same plate is shown before (top) and after (bottom) washing. (B) Pseudohyphal growth on SLAD media for diploid Sigma, S288c^{FLO8}, or Sigma/S288c^{FLO8} hybrids. (C) qPCR assay of *FLO11* transcript levels was performed on Sigma and S288c^{FLO8} strains that were WT or *tec1Δ*. Mean *FLO11* levels normalized to *ACT1* levels are presented ± SD. (*) P < 0.01 compared to WT.
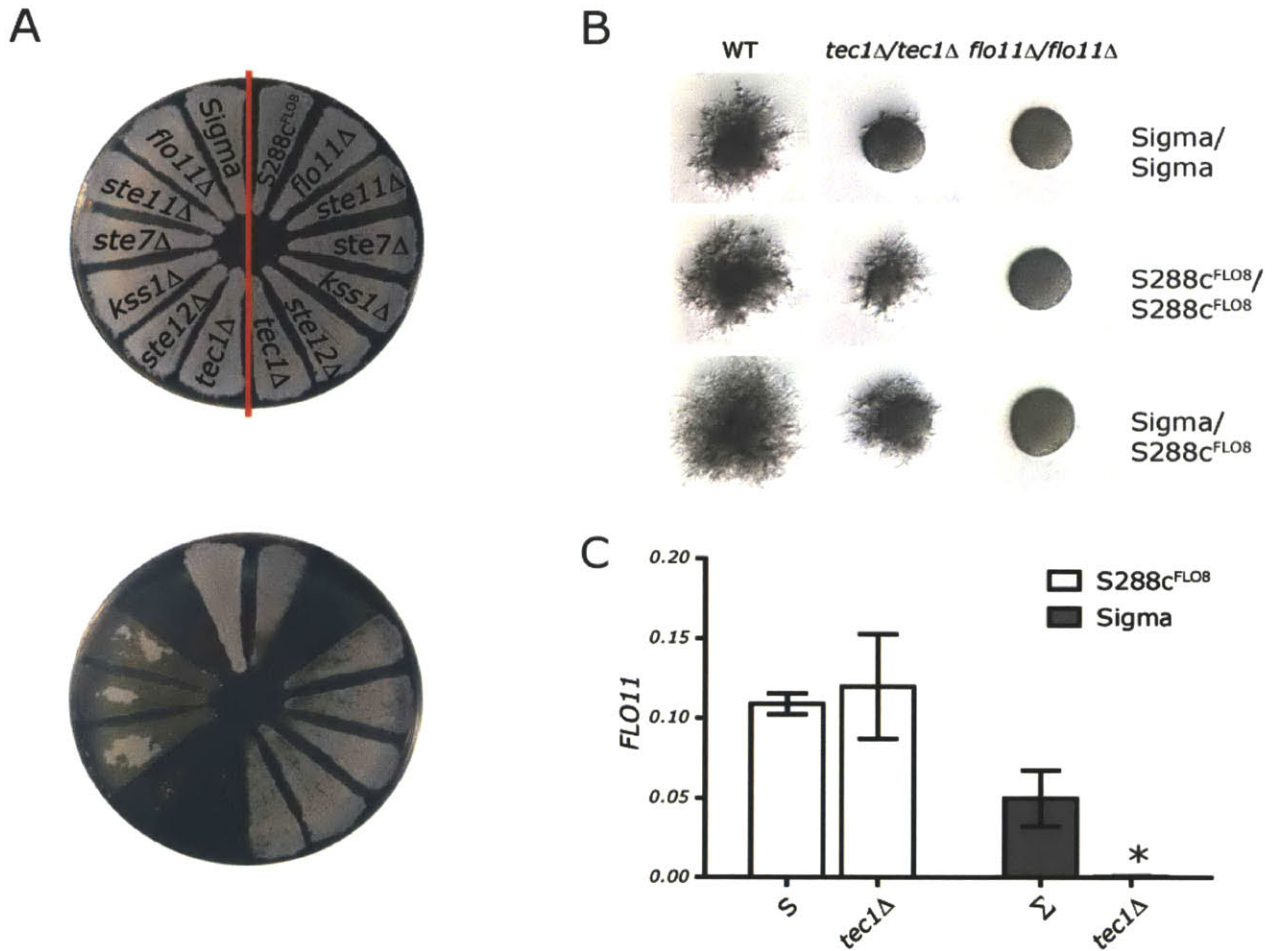
Figure 3-2 | **S288c with *FLO11pr^Sigma^::FLO11* is still fMAPK independent.** Agar adhesion assays performed on S288c^FLO8^ strains (right half of the plate), or Sigma strains (left half of the plate) in the promoter swap experiment (see text). The same plate is shown before (left) and after (right) washing. Strains with their endogenous *FLO11* promoter are labeled with their relevant genotype. Strains carrying a swapped *FLO11* promoter are labeled numerically: (1) S288c^FLO8^ *FLO11pr^Sigma^::FLO11*; (2) S288c^FLO8^ *FLO11pr^Sigma^::FLO11, tec1Δ*; (3) Sigma *FLO11pr^S288c^::FLO11*; (4) Sigma *FLO11pr^S288c^::FLO11, tec1Δ*.

Figure 3-3 | *tec1Δ* **bypass suppressors vary in the number of intragenic repeats.** The
S288c and Sigma alleles of *RPI1* and *MIT1* have intragenic repeats, but the repeat lengths differ
between the two strains. The schematic illustrates the alignment of the S288c and Sigma *RPI1*
and *MIT1*. The boxes represent individual repeat elements and arrow heads represent locations
of SNPs. Empty areas represent the shortened repeat length in that allele.
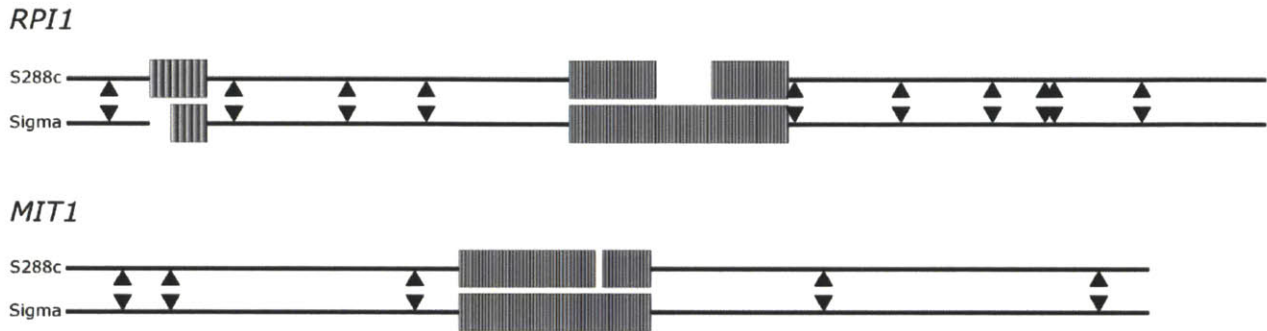
Figure 3-4 | *RPI1^{S288c}* can partially bypass the fMAPK pathway for agar adhesion and *FLO11* expression. (A) Agar adhesion of S288c^{FLO8} and Sigma strains carrying reciprocal allele swaps of *RPI1*. The top row shows adhesion assays performed on Sigma strains grown on synthetic media and the bottom row shows adhesion assays performed on S288c^{FLO8} strains grown on YPD (see METHODS). The same plates are shown before and after washing. qPCR assay of *FLO11* transcript levels performed on (B) S288c^{FLO8} strains grown in synthetic media and (C) Sigma strains grown on YPD. Mean *FLO11* levels normalized to *ACT1* levels are presented ± SD. ** P < 0.01. Strains with their endogenous *RPI1* allele are labeled with their relevant genotype. Strains carrying a swapped *RPI1* allele are labeled numerically: (1) S288c^{FLO8} *RPI1^{Sigma}*; (2) S288c^{FLO8} *RPI1^{Sigma}, tec1*Δ; (3) Sigma *RPI1^{S288c}*; (4) Sigma *RPI1^{S288c}, tec1*Δ.
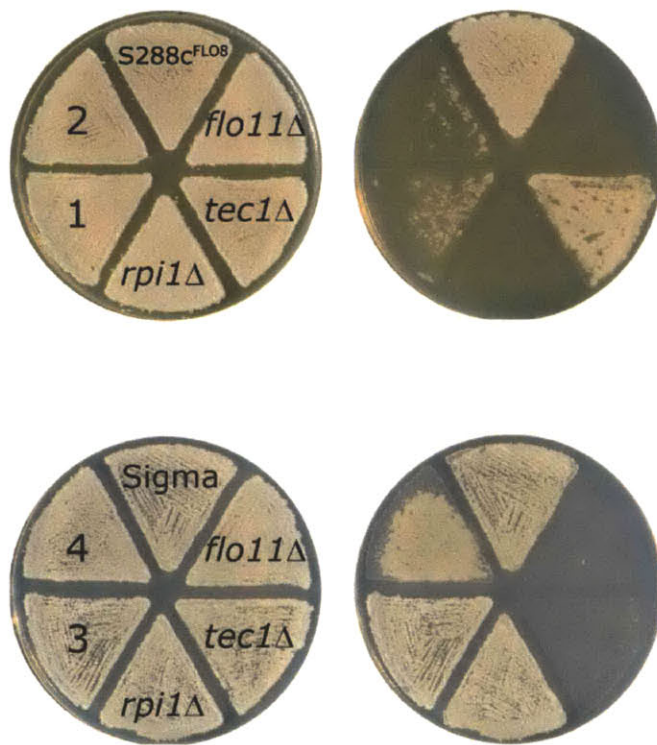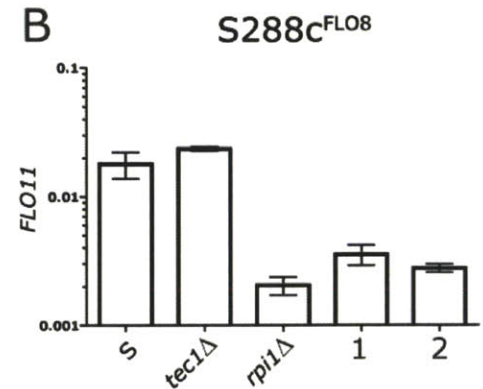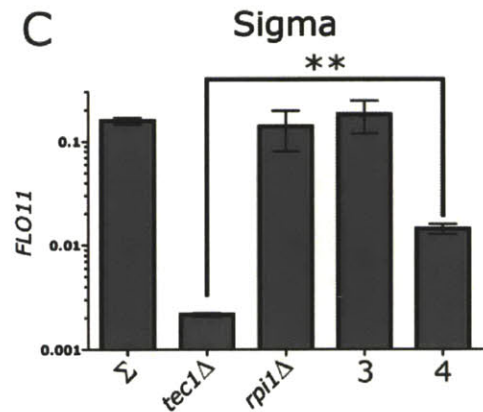
Figure 3-5 | **$RPI1^{S288c}$ shows strain independent localization to the *MIT1* and *FLO11***
promoters. Localization of Rpi1p using FLAG tagged alleles in Sigma and S288c^{FLO8} assayed
by ChIP followed by qPCR for enrichment at (A) -1.3kb in the *FLO11* promoter and (B) -1kb in
the *MIT1* promoter. Data were normalized to *ACT1* and are expressed as the mean fold
enrichment ± SD. * P < 0.01 compared to untagged.
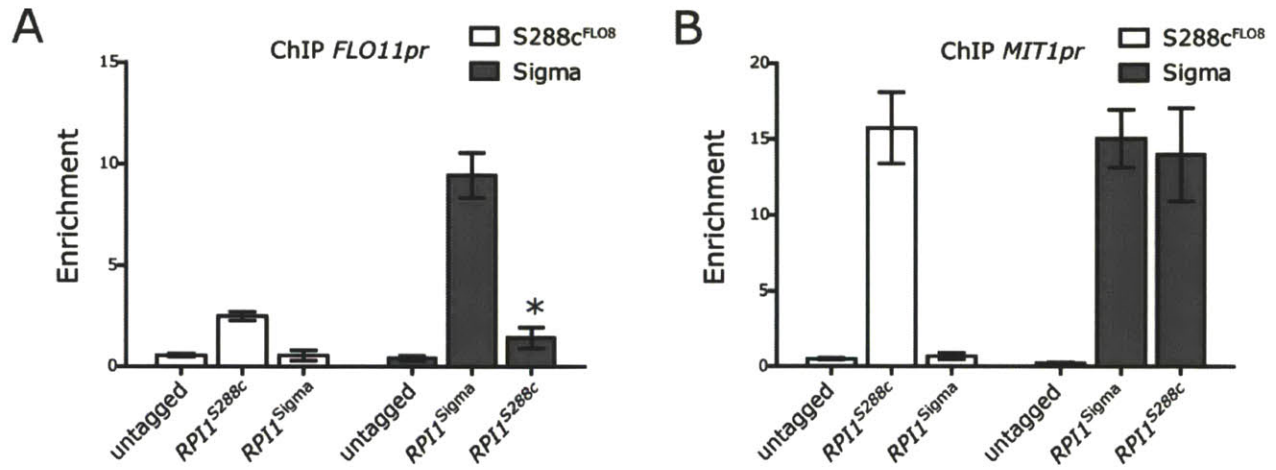
Figure 3-6 | **The Rpi1^{S288c} protein is hyperphosphorylated** Western blot analysis of RPI1 phosphorylation state in strains expressing either 3x flag tagged *RPI1*^{S288c} or *RPI1*^{Sigma}. Samples were treated with either buffer or lambda phosphatase.
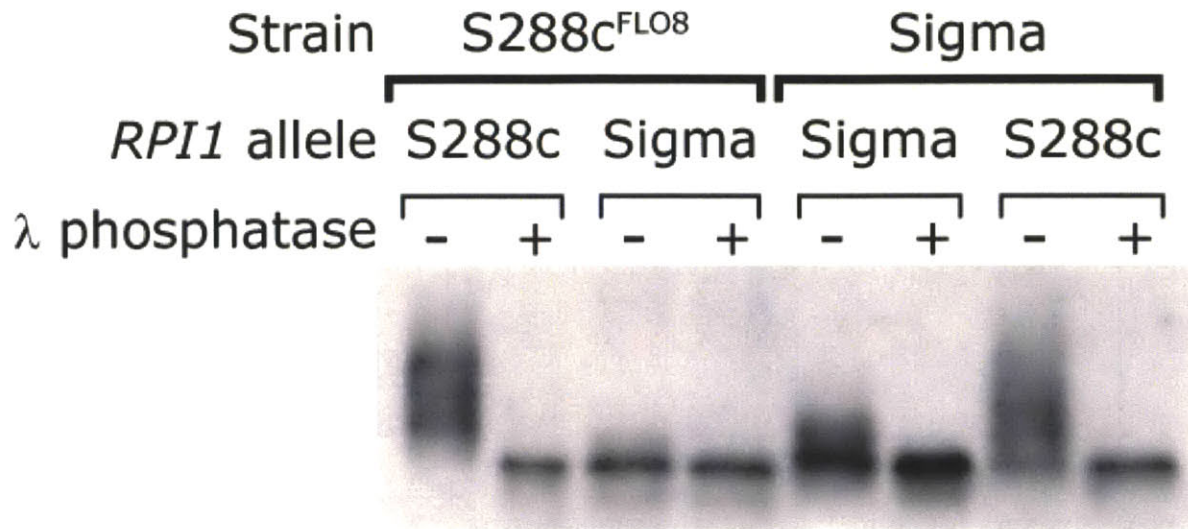
Figure 3-7 | **Many S288c genes differ from Sigma genes due to changes in intragenic tandem repeats.** 24 of the 107 genes predicted to differ between S288c and Sigma in the length of internal repeats were examined by PCR. 22 of these genes had the predicted size difference. Five genes are shown and the results for the other genes are shown in Supplemental Figure 4. *PGD1* and *SPT8* have two repeat regions that both change in size. For each pair the left sample is the S288c product and the right is the Sigma product.
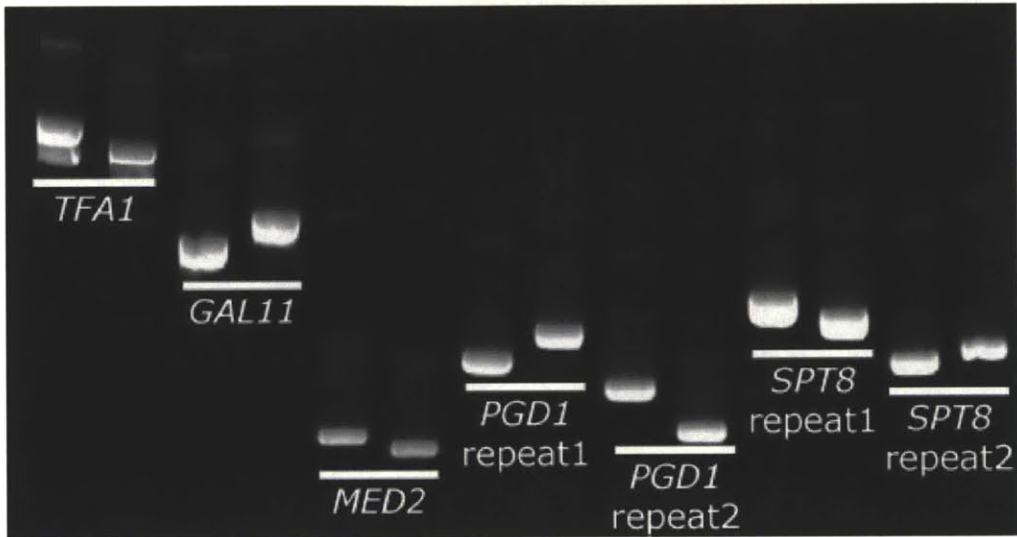
Supplemental Figure 3-1 │ **S288c with *FLO11pr^{Sigma} ::FLO11* is still fMAPK independent.** qPCR assay of *FLO11* transcript levels was performed on (A) S288c^{FLO8} and (B) Sigma strains carrying *FLO11* promoter swaps. Mean *FLO11* levels normalized to *ACT1* levels are presented ± SD. Strains with their endogenous *FLO11* promoter are labeled with their relevant genotype. Strains carrying a swapped *FLO11* promoter are labeled numerically: (1) S288c^{FLO8} *FLO11pr^{Sigma}::FLO11*; (2) S288c^{FLO8} *FLO11pr^{Sigma}::FLO11, tec1Δ*; (3) Sigma *FLO11pr^{S288c}::FLO11*; (4) Sigma *FLO11pr^{S288c}::FLO11, tec1Δ*.
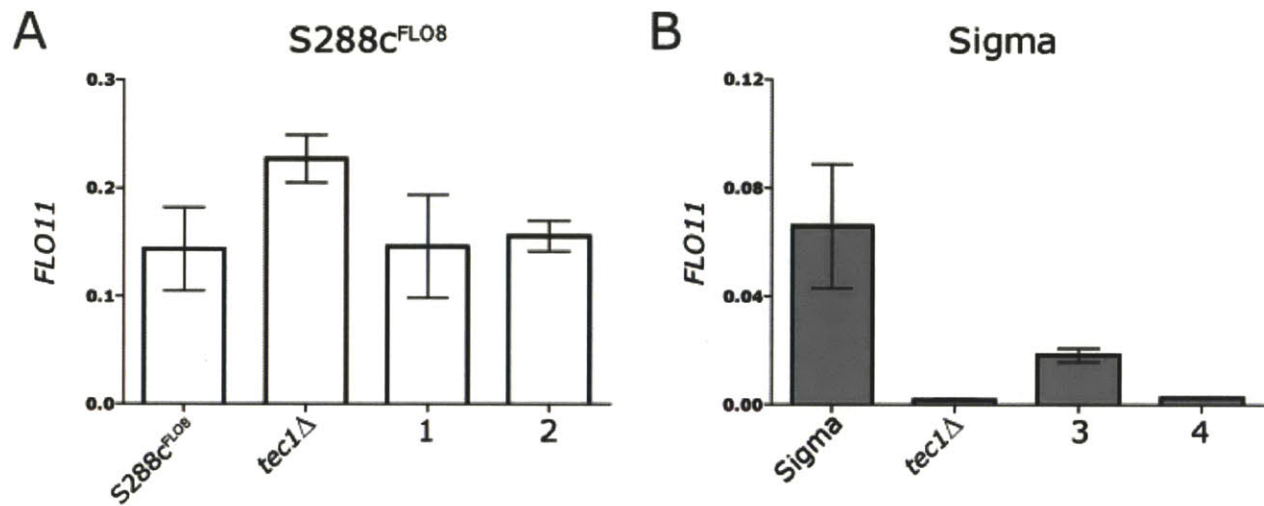
Supplemental Figure 3-2 | *tec1Δ* **bypass is a complex trait**. Agar adhesion assays of 24 tetrads from an S288c^FLO8 *tec1Δ* x Sigma *tec1Δ* cross. Two complete tetrads per row with one example underlined. Parental strains and controls spotted on the bottom of the plate.

Unwashed   ## Washed
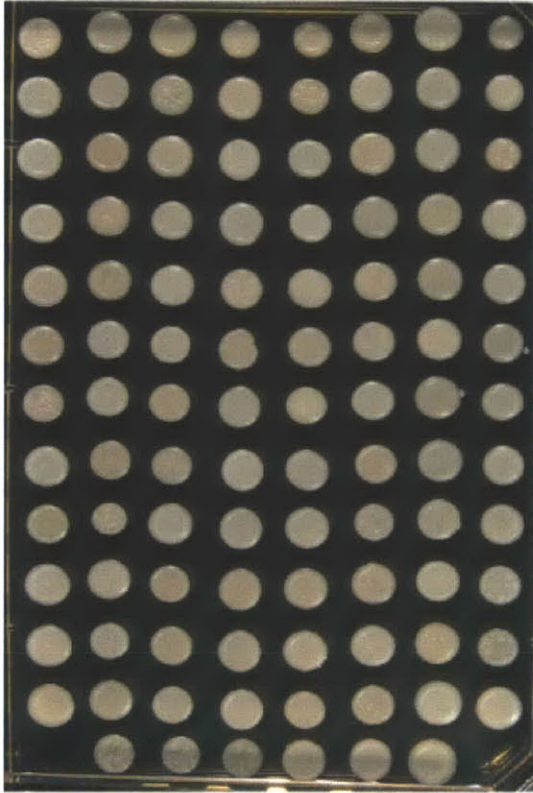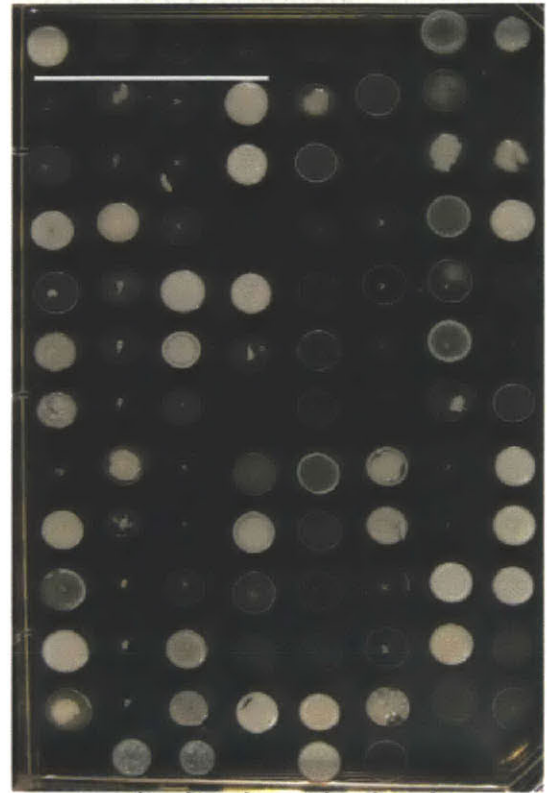
Supplemental Figure 3-3 | **fMAPK bypass of *FLO11* expression is a complex quantitative trait.** GFP fluorescence, measured in arbitrary units for (A) 276 F1 meiotic progney from a S288c^{FLO8} / Sigma *FLO11pr::GFP* / *FLO11pr::GFP tec1Δ* / *tec1Δ* diploid or (B) 276 meiotic progeny from the third generation of backcrossing (see methods). The average GFP fluorescence normalized to OD600 of 3 biological replicates are plotted. The progeny are sorted from highest to lowest fluorescence. Fluorescence of control strains are labeled and shown in green.
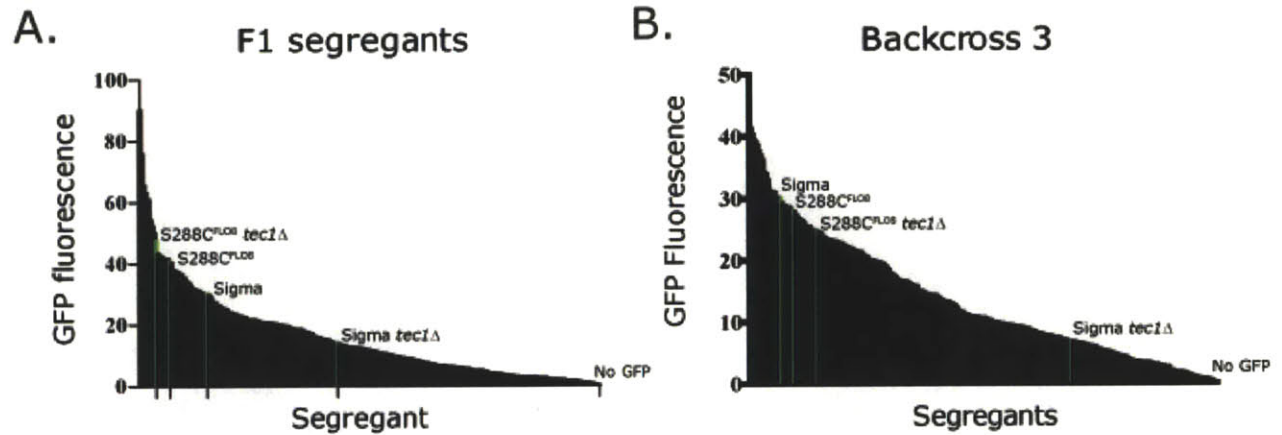
Supplemental Figure 3-4 | *RPI1* **and** *MIT1* **contain intragenic repeats**. Dot plot analysis of the S288c alleles of *RPI1* and *MIT1* compared against themselves. Repeat regions produce a characteristic box pattern

RPI1
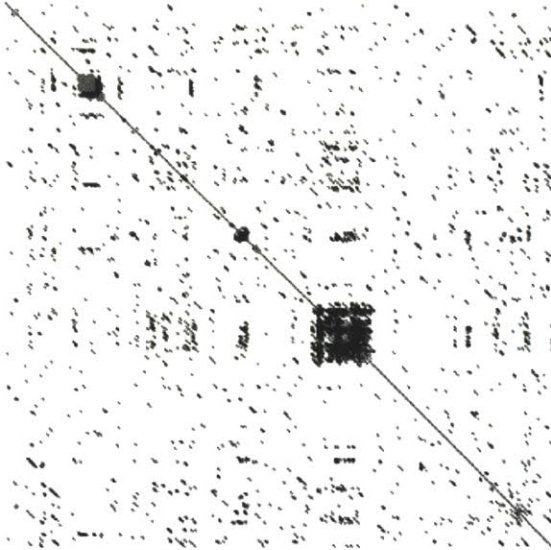
MIT1
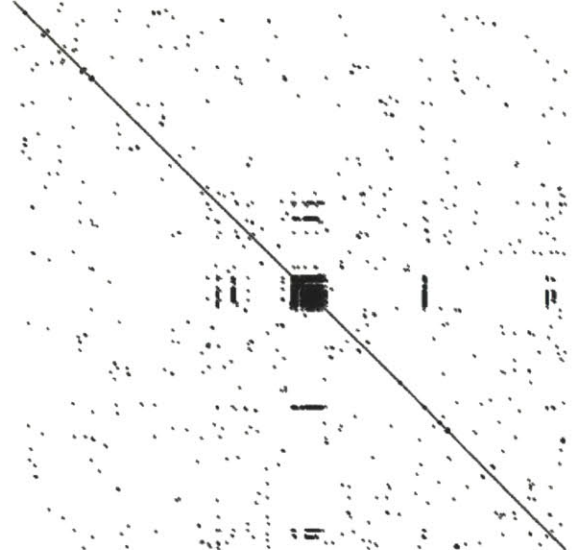
Supplemental Figure 3-5 | **Many genes have intragenic tandem repeats that differ in size between S288c and Sigma.** Four gels used to examine the length differences between S288c and Sigma for 24 genes and *FLO8* which was used as a control for a gene without repeats. 22/24 genes had the predicted repeat length differences. The genes *PGD1*, *SPT8*, and *SNF5* have two repeat regions that both changed in size. For each pair the left sample is S288c and the right sample is Sigma.
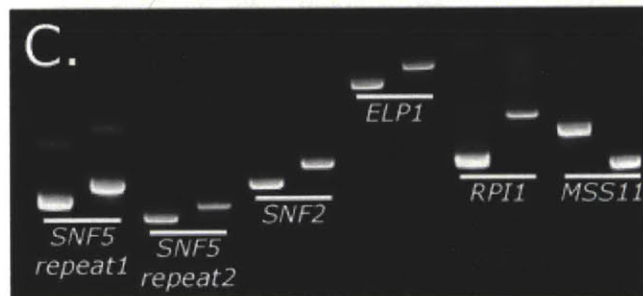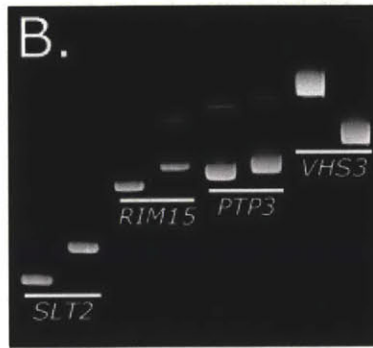
YAL054C	YIR020C	YOR021C	YDR485C
YNL020C	YJL218W	YOR029W	YML041C
YOR043W	YJR018W	YOR082C	YBR231C
YDR226W	YJR054W	YOR154W	YBR289W
YBL080C	YJR080C	YOR183W	YDR073W
YKR039W	YKL023W	YOR186W	YDR334W
YBR068C	YKL044W	YOR200W	YJL176C
YDR127W	YKL090W	YOR225W	YOR290C
YPR060C	YKL094W	YOR258W	YOL012C
YPR020W	YLL030C	YOR285W	YDL074C
YLR431C	YLL055W	YPL017C	YDR469W
YCR002C	YLR021W	YPL068C	YDR207C
YAR030C	YLR065C	YPL182C	YBR107C
YBL031W	YLR125W	YPL184C	YDR254W
YBL046W	YLR168C	YPL216W	YDR318W
YBR033W	YLR184W	YPL220W	YGR275W
YBR139W	YLR352W	YPL246C	YPR046W
YCL005W	YLR358C	YPL257W	YER068W
YCL036W	YLR374C	YPL260W	YAL012W
YCR016W	YLR434C	YPR170C	YER056C
YCR095C	YML010C-B	YER086W	YMR032W
YDL021W	YML010W-A	YDR200C	YNL166C
YDL073W	YMR135W-A	YCL058C	YNL229C
YDR003W	YMR158C-B	YBL006C	YLR420W
YDR248C	YMR191W	YPR030W	YML106W
YDR514C	YMR316C-A	YER083C	YJL115W
YER039C	YMR326C	YCR017C	YOL090W
YER048C	YNL023C	YGL027C	YLR418C
YER060W	YNL170W	YHR181W	YBR228W
YFL015C	YNL175C	YDL225W	YGL058W
YGL214W	YNL226W	YBR200W	YML021C
YGR071C	YNR025C	YHL003C	YOR144C
YHL017W	YOL032W	YLL026W	YDR364C
YHR080C	YOL042W	YJR060W	YCL061C
YHR210C	YOL048C	YDR176W	YMR048W
YIL059C	YOL159C	YGL066W	YBL082C
YIL086C		YLR055C	YKL213C
YIR014W		YNL107W	YDR069C

YDR320C	YMR125W	YLR024C	YMR174C
YNR006W	YBR034C	YGL203C	YHL019C
YJL095W	YDR432W	YPR087W	YBR053C
YKR054C	YDR195W	YER020W	
YBR159W	YGR019W	YML035C	
YBR171W	YPR101W	YBR221C	
YGR135W	YJR117W	YIL119C	
YFL011W	YPR049C	YKL109W	
YHR094C	YOL044W	YAL024C	
YBR133C	YGR004W	YER059W	
YOR178C	YNL173C	YPL219W	
YNL117W	YER053C	YMR179W	
YLR330W	YFL031W	YML014W	
YJL062W	YAL013W	YOL105C	
YDL035C	YDR174W	YOR008C	
YOR101W	YNR052C	YGL244W	
YKR029C	YKL043W	YHR087W	
YOL064C	YJL129C	YNR060W	
YGL045W	YDL230W	YBL075C	
YHL007C	YJL183W	YGR055W	
YOL101C	YKL139W	YGL033W	
YKR042W	YIL148W	YLR453C	
YOL091W	YGL236C	YGR104C	
YDL115C	YCL037C	YHR041C	
YLR219W	YDR500C	YPL144W	
YML128C	YHL033C	YPL258C	
YMR167W	YKL167C	YNL248C	
YMR031W-A	YLR185W	YJL189W	
YJR051W	YNL265C	YGR054W	
YLR180W	YOR096W	YNL125C	
YGR163W	YOR182C	YOR081C	
YAL047C	YPL090C	YPL212C	
YPL241C	YOR138C	YDR354W	
YLR368W	YHR034C	YKL211C	
YDR258C	YOR288C	YCL075W	
YNL076W	YMR091C	YDR330W	
YCL016C	YER110C	YHL016C	
YDR378C	YGL153W	YPR036W	
YKL009W	YIR004W	YLR373C	

Supplementary Table 3-2 | **Deletions leading to an Ahs⁻ phenotype only in Sigma.**

YKR024C	YGR162W	YPL031C	YIR021W
YHR114W	YAL048C	YDL044C	YER161C
YML022W	YPL259C	YBR191W	YGR123C
YLR278C	YLR370C	YGR105W	YDL069C
YGL258W	YNL271C	YKL119C	YDR197W
YGR271C-A	YMR267W	YOR085W	YML024W
YML117W	YDR079W	YNR051C	YBR165W
YOR267C	YDR529C	YEL059C-A	YER154W
YMR044W	YPL132W	YPL086C	YLR384C
YOR213C	YLR204W	YPL024W	YDR074W
YMR127C	YLR315W	YIL008W	YHL034C
YCR009C	YER156C	YFR019W	YDR096W
YCR088W	YLR375W	YPL193W	YDL081C
YIL034C	YFR048W	YJL124C	YOL023W
YMR008C	YGL188C-A	YPR040W	YIL125W
YGR040W	YGL211W	YDR512C	YDR120C
YGL014W	YGL228W	YNL098C	YGR020C
YDR005C	YKL037W	YOL051W	YOR332W
YNL053W	YOR141C	YDR289C	YFL054C
YOR002W	YKL110C	YGR257C	YGR272C
YOR067C	YDR276C	YLL041C	YBR026C
YDL159W	YBL007C	YNL037C	YHR011W
YGL019W	YBR245C	YOR136W	YCR105W
YGR188C	YGR062C	YEL051W	YPR116W
YLR362W	YLR337C	YKL080W	YCR079W
YHR021C	YLR056W	YDL067C	YER014C-A
YPR043W	YGR014W	YLR295C	YLR390W-A
YBR189W	YGR037C	YBL099W	YGR229C
YGR232W	YHL038C	YDR298C	YDR359C
YER118C	YGL252C	YBL066C	YLR385C
YMR312W	YAL002W	YBR162C	YOL068C
YPL101W	YOR334W	YLR404W	YMR263W
YKL143W	YOL115W	YNL097C	YCR077C
YDR184C	YGL003C	YGR180C	YHR120W
YDL190C	YPL005W	YCR086W	YER061C
YEL060C	YDR140W	YDR129C	YHR067W
YDL005C	YAL023C	YML008C	YBL071W-A
YGL025C	YDR477W	YGL084C	YER014W

YEL065W	YBR163W	YDR405W	YMR287C
YOR198C	YER087W	YER050C	YPL029W
YPL055C	YGL107C	YGR215W	YML055W
YDR393W	YGR102C	YHR147C	YLL033W
YHL020C	YMR066W	YHR168W	YMR228W
YGL246C	YMR098C	YIL093C	YJL102W
YER117W	YOR205C	YKL003C	YLR069C
YDL191W	YLR443W	YKL138C	YOR187W
YGL129C	YIL084C	YKL155C	YDR470C
YMR158W	YOR330C	YKL170W	YDR268W
YPL104W	YLR382C	YKR006C	YPL097W
YPR166C	YKL134C	YLR312W-A	YPL019C
YDR175C	YNL073W	YLR439W	YGR219W
YPL040C	YGR171C	YMR024W	YAL004W
YPL118W	YCR028C-A	YMR193W	
YLR192C	YDR296W	YNL005C	
YJL180C	YOL095C	YNL081C	
YER017C	YGL219C	YNL252C	
YMR089C	YNL213C	YPL173W	
YNL121C	YGR101W	YPR047W	
YPL148C	YLL006W	YBL090W	
YIL049W	YOL009C	YDR115W	
YNL119W	YOR211C	YDR337W	
YHR084W	YML062C	YEL050C	
YHR111W	YLR435W	YGL143C	
YIR019C	YDL090C	YGR165W	
YFL026W	YBR146W	YGR220C	
YNL180C	YBL038W	YHR091C	
YDR194C	YBR251W	YJL063C	
YKL149C	YBR268W	YKR085C	
YKL194C	YBR282W	YLR139C	
YPR163C	YCR003W	YMR097C	
YBR127C	YCR024C	YNL177C	
YMR293C	YCR046C	YOR150W	
YKL055C	YCR071C	YPR100W	
YOR221C	YDL045W-A	YPL002C	
YPL271W	YDR237W	YBL022C	
YDR332W	YDR322W	YBR083W	
YOR305W	YDR347W	YGL064C	

Supplementary Table 3-3 | **Deletions leading to an Ahs⁻ phenotype only in both S288c^{FLO8} and Sigma.**

YKL007W	YKR001C	YDR065W	YDR462W
YBR023C	YKL185W	YMR116C	YNR037C
YPL203W	YNL183C	YDL233W	YLR417W
YBL058W	YDR392W	YEL007W	YMR164C
YGR056W	YOR035C	YGR122W	YGR200C
YOL001W	YJL140W	YBR095C	YGR063C
YOL072W	YHR167W	YOR275C	YMR063W
YLR357W	YKL204W	YOR030W	YHL027W
YOL076W	YJR113C	YLR025W	YNL294C
YPL181W	YCL008C	YMR077C	YJL175W
YDR350C	YJR102C	YCR084C	
YMR154C	YOL004W	YDL006W	

Supplementary Table 3-4 | **ORFs with intragenic repeat length differences between S288c and Sigma.**

YAL035W	YOR053W	YJL123C
YAL064W-B	YOR156C	YJL130C
YBL011W	YOR290C	YJL162C
YBR017C	YPL049C	YKL028W
YBR030W	YPL229W	YKL032C
YBR212W	YPR142C	YKL105C
YCR067C	YPR143W	YKR092C
YDL005C	YPR152C	YKR102W
YDL035C	YAL065C	YLL010C
YDL122W	YAR050W	YLR055C
YDR133C	YBR289W	YLR106C
YDR134C	YCL043C	YLR114C
YDR232W	YDL037C	YLR177W
YDR273W	YDL039C	YLR406C-A
YDR299W	YDL058W	YML049C
YEL007W	YDR093W	YML113W
YFL024C	YDR150W	YMR016C
YFL033C	YDR420W	YMR044W
YGL013C	YDR517W	YMR124W
YGL237C	YER011W	YMR173W
YGR014W	YER030W	YMR173W-A
YHL020C	YER075C	YMR317W
YHR030C	YFL010C	YNL271C
YJL187C	YFL010W-A	YNL327W
YKL023W	YGL014W	YNR052C
YKL108W	YGR160W	YOR010C
YKL163W	YHL028W	YOR054C
YKR072C	YHR077C	YOR113W
YLL008W	YIL011W	YOR267C
YLR175W	YIL031W	YPL216W
YLR330W	YIL115C	YPR021C
YML074C	YIL119C	YPR123C
YMR070W	YIR010W	YPR124W
YMR136W	YIR019C	
YMR164C	YIR023W	
YNL186W	YJL020C	
YOL051W	YJL078C	

Supplementary Table 3-5 | **List of strains used in this study**

Strain	Genotype	Source
BY4741	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 flo8*-1	Brachmann et al. (1998)
yBC37	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 FLO8*	this study
yBC06A10	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 FLO8 tec1Δ*::KanMX4	this study
yBC06B5	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 FLO8 ste7Δ*::KanMX4	this study
yBC06G7	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 FLO8 ste11Δ*::KanMX4	this study
yBC07A3	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 FLO8 kss1Δ*::KanMX4	this study
yBC06B5	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 FLO8* ste12Δ::KanMX4	this study
yBC0192	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 flo11pr*^{S288c}*Δ*::*FLO11pr*^{Sigma} *FLO8*	this study
yBC0195	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 flo11pr*^{S288c}*Δ*::*FLO11pr*^{Sigma}*tec1Δ*::KanMX4 *FLO8*	this study
yBC11E2	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 flo11Δ*::GFP-*URA3 FLO8*	this study
yBC11H2	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 flo11Δ*::GFP-*URA3 tec1Δ*::KanMX4 *FLO8*	this study
yBC16A3	S288c MAT**a** *ura3Δ0 FLO8*	this study
yBC16F4	S288c MAT**a** /α *ura3Δ0/ura3Δ0 FLO8/FLO8*	this study
yBC20A1	S288c MAT**a** *ura3Δ0 tec1Δ*::hyg *FLO8*	this study
yBC20D1	S288c *MATα ura3Δ0 tec1Δ*::hyg *FLO8*	this study
yBC20A3	S288c MAT**a** /α *ura3Δ0/ura3Δ0 tec1Δ*::hyg/*tec1Δ*hyg *FLO8/FLO8*	this study
yBC11E8	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 flo11Δ*::HIS3PEST *FLO8*	this study
yBC11H8	S288c MAT**a** *his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 flo11Δ*::HIS3PEST *tec1Δ*::KanMX4 *FLO8*	this study
yBC18A1	S288c MAT**a** *ura3Δ0 rpi1Δ*::*URA3 FLO8*	this study
yBC18A6	S288c MAT**a** *ura3Δ0 rpi1Δ*::*RPI1*^{Sigma} *FLO8*	this study
yBC18A8	S288c MAT**a** *ura3Δ0 rpi1Δ*::*RPI1*^{Sigma} tecΔ1::KanMX4 *FLO8*	this study
yBC29A9	S288c *MATa ura3Δ0 RPI1*-3xFLAG-*URA3 FLO8*	this study

yBC29D9	S288c MAT**a** *ura3Δ0 rpi1Δ::RPI1*^{Sigma}-3xFLAG-*URA3 FLO8*	this study
10560-6B	Sigma *MATα his3*::hisG *leu2*::hisG *trp1*::hisG *ura3-52*	Fink Collection
yBC0172	Sigma MAT**a** *his3*::hisG *leu2*::hisG *trp1*::hisG *ura3-52*	this study
Sigma *tec1Δ*	MAT**a** *can1Δ::STE2pr-Sphis5 lyp1Δ::STE3pr-LEU2 his3*::hisG *leu2Δ ura3Δ tec1Δ*::KanMX4	Dowell and Ryan et al. (2010)
Sigma *ste7Δ*	MAT**a** *can1Δ::STE2pr-Sphis5 lyp1Δ::STE3pr-LEU2 his3*::hisG *leu2Δ ura3Δ ste7Δ*::KanMX4	Dowell and Ryan et al. (2010)
Sigma *ste11Δ*	MAT**a** *can1Δ::STE2pr-Sphis5 lyp1Δ::STE3pr-LEU2 his3*::hisG *leu2Δ ura3Δ ste11Δ*::KanMX4	Dowell and Ryan et al. (2010)
Sigma *kss1Δ*	MAT**a** *can1Δ::STE2pr-Sphis5 lyp1Δ::STE3pr-LEU2 his3*::hisG *leu2Δ ura3Δ kss1Δ*::KanMX4	Dowell and Ryan et al. (2010)
Sigma *ste12Δ*	MAT**a** *can1Δ::STE2pr-Sphis5 lyp1Δ::STE3pr-LEU2 his3*::hisG *leu2Δ ura3Δ ste12Δ*::KanMX4	Dowell and Ryan et al. (2010)
yBC0193	Sigma MAT**a** *his3*::hisG *leu2*::hisG *trp1*::hisG *ura3-52 flo11*pr^{Sigma}Δ*::FLO11*pr^{S288c}	this study
yBC0196	Sigma MAT**a** *his3*::hisG *leu2*::hisG *trp1*::hisG *ura3-52 flo11*pr^{Sigma}Δ*::FLO11*pr^{S288c} *tec1Δ*::KanMX4	this study
yBC11G1	Sigma MAT**a** *his3*::hisG *leu2*::hisG *trp1*::hisG *ura3-52 flo11Δ*::GFP-*URA3*	this study
yBC11B2	Sigma MAT**a** *his3*::hisG *leu2*::hisG *trp1*::hisG *ura3-52 flo11Δ*::GFP-*URA3 tec1Δ*::KanMX4	this study
yBC16H3	Sigma MAT**a** *ura3-52*	this study
yBC16B4	Sigma *MATα ura3-52*	this study
yBC16G4	Sigma MAT**a**/α *ura3-52/ura3-52*	this study
yBC20G1	Sigma MAT**a** *ura3-52 tec1Δ*::hyg	this study
yBC20B2	Sigma *MATα ura3-52 tec1Δ*::hyg	this study
yBC20C3	Sigma MAT**a**/α *ura3-52/ura3-52 tec1Δ*::hyg/*tec1Δ*hyg *FLO8/FLO8*	this study
yBC11A7	Sigma MAT**a** *his3*::hisG *leu2*::hisG *trp1*::hisG *ura3-52 flo11Δ::HIS3*PEST	this study
yBC11D7	Sigma MAT**a** *his3*::hisG *leu2*::hisG *trp1*::hisG *ura3-52 flo11Δ::HIS3*-PEST *tec1Δ*::KanMX4	this study
yBC18G1	Sigma MAT**a** *ura3-52 rpi1Δ::URA3*	this study
yBC18G6	Sigma MAT**a** *ura3-52 rpi1Δ::RPI1*^{S288c}	this study
yBC18G8	Sigma MAT**a** *ura3-52 rpi1Δ::RPI1*^{Sigma} *tecΔ1*::KanMX4	this study
yBC29G9	Sigma MAT**a** *ura3-52 RPI1*-3xFLAG-*URA3*	this study

yBC29B10	Sigma MAT**a** *ura3-52 rpi1*Δ*::RPI1*^{Sigma}-3xFLAG-*URA3*	this study
yBC09H1	S288c^{FLO8}/Sigma MAT**a** */α ura3*Δ*0/ura3-52 his3*Δ*0/his3*::hisG *leu2*Δ*0/leu2*::hisG *met15*Δ*0/MET15 TRP1/trp1*::hisG *tec1*Δ::hyg/*tec1*Δ::hyg *flo11*Δ::GFP-*URA3/flo11*Δ::GFP-*URA3*	this study
yBC03A10	S288c^{FLO8}/Sigma MAT**a** */α ura3*Δ*0/ura3-52 his3*Δ*0/his3*::hisG *met15*Δ*0/MET15 tec1*Δ::KanMX4/*tec1*Δ::KanMX	this study

Supplementary Table 3-6 | **List of oligonucleotides used in this study**

Name	Sequence (5' to 3')	Description
BCP10	agtgcttaaccggaacaaacc	*FLO8F*
BCP15	tatgatcatgatttacgatgaccgt	*FLO8R*
BCP46	ggaaacaagctgagctggac	Flanking *TEC1*
BCP47	tcgtggtttcatccaagtga	Flanking *TEC1*
BCP191	cccaagcgagacctagagtg	Flanking *STE12*
BCP192	gaacatcgatgccttcacct	Flanking *STE12*
BCP195	aagtgattcgtggggtaacg	Flanking *STE7*
BCP196	tgggttattaatcgccttcg	Flanking *STE7*
BCP199	attctcgcccaacttttcct	Flanking *STE11*
BCP200	tcttcgtgcttccatctgtg	Flanking *STE11*
BCP236	tccccttggtgaaagaaatg	Flanking *kss1*
BCP237	ttgattacagtcgcgtcagc	Flanking *kss1*
BCP249	GGTTCTAATTAAAATATACTTTTGTAGGCCTCAA AAATCCATATACGCACACTatgacagagcagaaagccctag	to replace the *FLO11* ORF with *HIS3*
BCP257	tgatgagggtgaagggaaac	*RPI1* swap
BCP316	ggtGCATCCAACTTGAACATTTCGAGAAAGC	For amplifying PEST seq from *CLN2*
BCP317	CTATATTACTTGGGTATTGCCCATACC	For amplifying PEST seq from *CLN2*
BCP320	GCTTTCTCGAAATGTTCAAGTTGGATGCacccataaga acacctttggtggag	linearize pRS313 to add PEST seq from *CLN2*
BCP321	GGTATGGGCAATACCCAAGTAATATAGtgacaccgatt atttaaagctg	linearize pRS313 to add PEST seq from *CLN2*
BCP324	atttaagaatgaaaacatcgtaatgaagaaacgaacatgttggaattgtatcaCT ATATTACTTGGGTATTGCCCATACC	To replace *FLO11* with *HIS3PEST*
BCP358	CTTTTTTTTAAGTCTTTTTTTTTTTTTCTCATCATTT TATTACTGATATTTATAAAagattgtactgagagtgcac	*rpi1::ura3*
BCP359	TAGAATTAAAGGGGTAGAAAATTTATGGTGGAG ACTTCCCGATACATACTctgtgcggtatttcacaccg	*rpi1::ura3*
BCP360	cgtattcgtttaactatttctcagtcc	*RPI1* swap
BCP412	ctcaacagcagatccagcag	*MSS11F* repeats
BCP413	gaaggcataagtccggttga	*MSS11R* repeats
BCP419	cattgaagccgaacaagaatg	*RPI1F* repeats
BCP420	cttgactgaatatgctctggtg	*RPI1R* repeats
BCP423	tgcaagatttcaggctgttt	*SLT2F* repeats
BCP424	atccacatctgaaggctgct	*SLT2R* repeats

BCP534	GACTACAAGGATGATGACGATAAAGGTGACTAT AAAGATCATGACATTGATTATAAAGACCATGACT AAgcaggtcgacaacccttaat	to build a C terminal flag tagging construct
BCP535	GCGGCCGCATAGGCCACT	to build a C terminal flag tagging construct
BCP536	ACCGTTGCATAATATGTCAACTTCAGACTCAGAA AATTTTATGCAACAACATgactacaaggatgatgacgata	C-terminally tag *RPI1* with FLAG
BCP537	GAATTAAAGGGGTAGAAAATTTATGGTGGAGAC TTCCCGATACATACTTTAgcggccgcataggccact	C-terminally tag *RPI1* with FLAG
BCP572	cattaaacccgtggaacagc	*GAL11F* repeats
BCP573	gggaataggtgccactttca	*GAL11R* repeats
BCP574	ctgaatgggtggatccaaat	*URA2F* repeats
BCP575	agaacagatggatcacctgga	*URA2R* repeats
BCP576	gaaccggcaagacttaacca	*EPL1F* repeats
BCP577	ttctgtttcgcttctgaattg	*EPL1R* repeats
BCP580	ggacaggagcaggaagaaaa	*NUP159F* repeats
BCP581	tccgaatgcagatgtaccaa	*NUP159R* repeats
BCP584	atgggcataaacggtgacat	*VHS3F* repeats
BCP585	agatcgctgtagccctcctt	*VHS3R* repeats
BCP586	aacctgcacaggaaacatcc	*TFA1F* repeats
BCP587	ctgaagcagtggcagtagca	*TFA1R* repeats
BCP588	cccacgactacaagcacaaa	*WSC4F* repeats
BCP589	cttgtagaaatgggggctga	*WSC4R* repeats
BCP628	aaggctgcagtggtcaagtt	*DNF2F* repeats
BCP629	atatctgaactgcccgatgg	*DNF2R* repeats
BCP632	tacaatcccacgcagtttca	*ULP2F* repeats
BCP633	ttccgtagttgcatcatcaaa	*ULP2R* repeats
BCP634	gctggaaaacgactcaaagc	*SPT8F* repeats
BCP635	agcagccttttgctcatcat	*SPT8R* repeats
BCP636	atgatgagcaaaaggctgct	*SPT8F* repeats
BCP637	tccattagcagaggcttcgt	*SPT8R* repeats
BCP638	ctgtgtcaggacgccataga	*RIM15F* repeats
BCP639	tccttggggaaaactgaaaa	*RIM15R* repeats
BCP640	tcaaatgtgatgccaggttc	*SNF2F* repeats
BCP641	ttgctcggcagtaaacattg	*SNF2R* repeats
BCP642	agtacggggaccttgaacct	*SWE1F* repeats
BCP643	tacgagaatccacgctttcc	*SWE1R* repeats
BCP644	cagctggtgttcagggaaat	*PTP3F* repeats
BCP645	ccaaatcaggccaatttttc	*PTP3R* repeats

111

BCP646	acaacggcgatgaaaagaat	*MED2F* repeats
BCP647	tgccgttatcgtcattgttg	*MED2R* repeats
BCP648	aggctggataacctgcaaga	*DSN1F* repeats
BCP649	ttgcagtcgcatctccacta	*DSN1R* repeats
BCP650	caagaccattcgctgcagta	*IXR1F* repeats
BCP651	taaggcgcttgttgttgttg	*IXR1R* repeats
BCP654	atgggaactccaaccgtaca	*PGD1F* repeats
BCP655	agtcgactgctgtgcgtaga	*PGD1R* repeats
BCP656	ccaataacaccccgctacag	*PGD1F* repeats
BCP657	tactgtggttgaggctgctg	*PGD1R* repeats
BCP658	tagtttgaaggaacgcgaca	*UBP10F* repeats
BCP659	gaacccaagttttcaccaatg	*UBP10R* repeats
BCP660	atgattcagcaacgacacca	*SNF5F* repeats
BCP661	aggaggaggggtagaagtcg	*SNF5R* repeats
BCP662	tgttgcacaacaacaagtgc	*SNF5F* repeats
BCP663	gctgttgtcgctgtatttgg	*SNF5R* repeats
FLO11 FW	cacttttgaagtttatgccacacaag	*FLO11* qPCR
FLO11 RV	cttgcatattgagcggcactac	*FLO11* qPCR
ACT1 FW	ctccaccactgctgaaagagaa	*ACT1* qPCR
ACT1 RV	ccaaggcgacgtaacatagtttt	*ACT1* qPCR

References

BORNEMAN, A. R., T. A. GIANOULIS, Z. D. ZHANG, H. YU, J. ROZOWSKY *et al.*, 2007a Divergence of transcription factor binding sites across related yeast species. Science **317:** 815-819.

BORNEMAN, A. R., J. A. LEIGH-BELL, H. YU, P. BERTONE, M. GERSTEIN *et al.*, 2006 Target hub proteins serve as master regulators of development in yeast. Genes Dev **20:** 435-448.

BORNEMAN, A. R., Z. D. ZHANG, J. ROZOWSKY, M. R. SERINGHAUS, M. GERSTEIN *et al.*, 2007b Transcription factor binding site identification in yeast: a comparison of high-density oligonucleotide and PCR-based microarray platforms. Funct Integr Genomics **7:** 335-345.

BREM, R. B., G. YVERT, R. CLINTON and L. KRUGLYAK, 2002 Genetic dissection of transcriptional regulation in budding yeast. Science **296:** 752-755.

BROWN, T., 2001 Southern Blotting in *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc.

CAIN, C. W., M. B. LOHSE, O. R. HOMANN and A. D. JOHNSON, 2011 A Conserved Transcriptional Regulator Governs Fungal Morphology in Widely Diverged Species. Genetics.

CARLBORG, O., and C. S. HALEY, 2004 Epistasis: too often neglected in complex trait studies? Nat Rev Genet **5:** 618-625.

DEMOGINES, A., E. SMITH, L. KRUGLYAK and E. ALANI, 2008 Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. PLoS Genet **4:** e1000123.

DICKSON, S. P., K. WANG, I. KRANTZ, H. HAKONARSON and D. B. GOLDSTEIN, 2010 Rare variants create synthetic genome-wide associations. PLoS Biol **8:** e1000294.

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

ELION, E. A., J. A. BRILL and G. R. FINK, 1991 FUS3 represses CLN1 and CLN2 and in concert with KSS1 promotes signal transduction. Proc Natl Acad Sci U S A **88:** 9392-9396.

FIDALGO, M., R. R. BARRALES and J. JIMENEZ, 2008 Coding repeat instability in the FLO11 gene of Saccharomyces yeasts. Yeast **25:** 879-889.

GATBONTON, T., M. IMBESI, M. NELSON, J. M. AKEY, D. M. RUDERFER *et al.*, 2006 Telomere length as a quantitative trait: genome-wide survey and genetic mapping of telomere length-control genes in yeast. PLoS Genet **2:** e35.

GIMENO, C. J., P. O. LJUNGDAHL, C. A. STYLES and G. R. FINK, 1992 Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. Cell **68:** 1077-1090.

GUTHRIE, C., and G. FINK, 2002 *Guide to yeast genetics and molecular and cellular biology*.

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

JAKOBSDOTTIR, J., M. B. GORIN, Y. P. CONLEY, R. E. FERRELL and D. E. WEEKS, 2009 Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers. PLoS Genet **5:** e1000337.

KELLIS, M., N. PATTERSON, M. ENDRIZZI, B. BIRREN and E. S. LANDER, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423:** 241-254.

KORBEL, J. O., A. E. URBAN, J. P. AFFOURTIT, B. GODWIN, F. GRUBERT *et al.*, 2007 Paired-end mapping reveals extensive structural variation in the human genome. Science **318:** 420-426.

LANE, S., S. ZHOU, T. PAN, Q. DAI and H. LIU, 2001 The basic helix-loop-helix transcription factor Cph2 regulates hyphal development in Candida albicans partly via TEC1. Mol Cell Biol **21:** 6418-6428.

LEVDANSKY, E., J. ROMANO, Y. SHADKCHAN, H. SHARON, K. J. VERSTREPEN *et al.*, 2007 Coding tandem repeats generate diversity in Aspergillus fumigatus genes. Eukaryot Cell **6:** 1380-1391.

LITI, G., D. M. CARTER, A. M. MOSES, J. WARRINGER, L. PARTS *et al.*, 2009 Population genomics of domestic and wild yeasts. Nature **458:** 337-341.

LIU, H., C. A. STYLES and G. R. FINK, 1993 Elements of the yeast pheromone response pathway required for filamentous growth of diploids. Science **262:** 1741-1744.

113

LIU, H., C. A. STYLES and G. R. FINK, 1996 Saccharomyces cerevisiae S288C has a mutation in FLO8, a gene required for filamentous growth. Genetics **144**: 967-978.

LO, W. S., and A. M. DRANGINIS, 1998 The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by Saccharomyces cerevisiae. Mol Biol Cell **9**: 161-171.

LORENZ, M. C., and J. HEITMAN, 1998 Regulators of pseudohyphal differentiation in Saccharomyces cerevisiae identified through multicopy suppressor analysis in ammonium permease mutant strains. Genetics **150**: 1443-1457.

MACDONALD, M. E., C. M. AMBROSE, M. P. DUYAO, R. H. MYERS, C. LIN et al., 1993 A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell **72**: 971-983.

MANOLIO, T. A., F. S. COLLINS, N. J. COX, D. B. GOLDSTEIN, L. A. HINDORFF et al., 2009 Finding the missing heritability of complex diseases. Nature **461**: 747-753.

MONTEIRO, P. T., N. D. MENDES, M. C. TEIXEIRA, S. D'OREY, S. TENREIRO et al., 2008 YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in Saccharomyces cerevisiae. Nucleic Acids Res **36**: D132-136.

NOGAMI, S., Y. OHYA and G. YVERT, 2007 Genetic complexity and quantitative trait loci mapping of yeast morphological traits. PLoS Genet **3**: e31.

RICE, P., I. LONGDEN and A. BLEASBY, 2000 EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16**: 276-277.

ROBERTS, R. L., and G. R. FINK, 1994 Elements of a single MAP kinase cascade in Saccharomyces cerevisiae mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev **8**: 2974-2985.

ROSE, M. D., P. NOVICK, J. H. THOMAS, D. BOTSTEIN and G. R. FINK, 1987 A Saccharomyces cerevisiae genomic plasmid bank based on a centromere-containing shuttle vector. Gene **60**: 237-243.

ROZEN, S., and H. SKALETSKY, 2000 Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol **132**: 365-386.

SCHACHERER, J., J. A. SHAPIRO, D. M. RUDERFER and L. KRUGLYAK, 2009 Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. Nature **458**: 342-345.

SHEETS, A. J., and J. W. ST GEME, 3RD, 2011 Adhesive activity of the haemophilus cryptic genospecies cha autotransporter is modulated by variation in tandem Peptide repeats. J Bacteriol **193**: 329-339.

TAN, J. C., A. TAN, L. CHECKLEY, C. M. HONSA and M. T. FERDIG, 2010 Variable numbers of tandem repeats in Plasmodium falciparum genes. J Mol Evol **71**: 268-278.

TSONG, A. E., M. G. MILLER, R. M. RAISNER and A. D. JOHNSON, 2003 Evolution of a combinatorial transcriptional circuit: a case study in yeasts. Cell **115**: 389-399.

TSONG, A. E., B. B. TUCH, H. LI and A. D. JOHNSON, 2006 Evolution of alternative transcriptional circuits with identical logic. Nature **443**: 415-420.

VERSTREPEN, K. J., A. JANSEN, F. LEWITTER and G. R. FINK, 2005 Intragenic tandem repeats generate functional variability. Nat Genet **37**: 986-990.

VOYNOV, V., K. J. VERSTREPEN, A. JANSEN, V. M. RUNNER, S. BURATOWSKI et al., 2006 Genes with internal repeats require the THO complex for transcription. Proc Natl Acad Sci U S A **103**: 14423-14428.

WANG, H., D. MAYHEW, X. CHEN, M. JOHNSTON and R. D. MITRA, 2011 Calling Cards enable multiplexed identification of the genomic targets of DNA-binding proteins. Genome Res **21**: 748-755.

WINZELER, E. A., H. LIANG, D. D. SHOEMAKER and R. W. DAVIS, 2000 Functional analysis of the yeast genome by precise deletion and parallel phenotypic characterization. Novartis Found Symp **229**: 105-109; discussion 109-111.

WINZELER, E. A., D. D. SHOEMAKER, A. ASTROMOFF, H. LIANG, K. ANDERSON *et al.*, 1999 Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science **285:** 901-906.

ZEITLINGER, J., I. SIMON, C. T. HARBISON, N. M. HANNETT, T. L. VOLKERT *et al.*, 2003 Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. Cell **113:** 395-404.

Chapter 4

Summary, Discussion, and Future Directions

Summary

The work presented in this thesis described the genomic and phenotypic comparison of two closely related strains, Σ1278b (Sigma) and S288c. 49% of the ORFs were found to be completely identical between these two strains and most of the variation is present in the subtelomeric regions (DOWELL *et al.* 2010). Also, no large translocations or inversions exist that prevent the two strains from mating and producing viable progeny at high frequency. Despite these findings, systematic analysis of deletion strains revealed that the two strains differ significantly in the functions required for life and agar adhesion.

In Chapter Two, I presented the high quality assembly of the Sigma genome. By combining 7x coverage of the genome from Sanger sequencing, 60x coverage of the genome from Illumina deep sequencing, and experimental validation of novel features in the sequence, we obtained a genome sequence for Sigma that is currently of higher quality than of any other *S. cerevisiae* strain besides the S288c reference genome (KELLIS *et al.* 2003; LITI *et al.* 2009; SCHACHERER *et al.* 2009).

The high quality Sigma genome sequence was then used to guide the construction of a deletion library in the Sigma background. This library is a collection of deletion mutants such that every non-essential gene in Sigma is deleted. This deletion library is comparable to the deletion library made in the S288c background and allows a direct comparison of mutant phenotypes between two the strains (DOWELL *et al.* 2010; WINZELER *et al.* 2000). A comparison of the essential genes in each strain revealed that 6% of these genes are essential in one strain but

not the other. For the majority of the strain specific essential genes a cross between S288c and

Sigma will segregate ≥5 modifiers that allow for bypass of the lethality of these strain specific

essentials. This genetic complexity precluded the identification of any one of the bypassers.

In Chapter Three I presented screens of the deletion library that examined the functions

required for adhesion in S288c^{FLO8} and Sigma. These screens revealed many functions had strain

specific roles in adhesion. The filamentation MAPK (fMAPK) pathway is a prominent example

of these strain specific adhesion functions because it is only required for adhesion in the Sigma

strain. Subsequent selections revealed that the S288c allele of the *RPI1* gene could bypass the

fMAPK pathway for *FLO11* expression. Allele swaps of *RPI1* showed that *RPI1*^{S288c} can bypass

the fMAPK for *FLO11* transcription in Sigma and it can interact with the *FLO11* promoter in

both strains. While *RPI1*^{Sigma} can also interact with the *FLO11* promoter, it can only do so in the

Sigma strain and it has no effect on *FLO11* transcription in either strain. It is possible that Rpi1p

can bind to the *FLO11* promoter but be inactive. This possibility is supported by a difference in

the levels of phosphorylation of the two alleles. The phosphorylation of Rpi1p^{S288c} is

independent of strain background, but the phosphorylation of Rpi1p^{Sigma} is dependent upon a

factor specific to the Sigma background.

There are numerous sequence differences between *RPI1*^{S288c} and *RPI1*^{Sigma} that could

account for the allele specific activities. While the S288c and Sigma alleles of *RPI1* differ by ten

SNPs, they also differ in the size of two intragenic repeat regions. While previous studies have

shown that tracks of tandem repeats within a gene can differ in size, these size changes were

mostly thought to be restricted to cell surface genes (LEVDANSKY *et al.* 2007; SHEETS and ST

GEME 2011; TAN *et al.* 2010; VERSTREPEN *et al.* 2005). The change in size of a transcription

factor led us to examine if changes in gene size occurs more frequently than previously expected.

Indeed, over 100 genes changed in size due to a change in the number of tandem repeats. These genes are varied in function and include cell wall genes, chromatin modifiers, and cell cycle regulators.

Discussion and future directions

Linkage mapping and deletion libraries

Studies attempting to uncover natural variation via linkage mapping often obtain five or fewer loci significantly associated with a trait. Additionally, the causal variants in these loci are frequently not obvious, and therefore it is not clear how these loci are affecting the trait (BEN-ARI et al. 2006; DEMOGINES et al. 2008; DEUTSCHBAUER and DAVIS 2005; NOGAMI et al. 2007; SINHA et al. 2008; STEINMETZ et al. 2002). In contrast to linkage studies, our study of natural variation using deletion libraries has uncovered hundreds of genes that preferentially affect agar adhesion in one strain or the other, and many of these genes have a proposed role in the cell. This large set of genes has provided insight into which pathways each strain utilizes for agar adhesion. But how can so many genes have strain specific effects? One possible explanation is that examining the deletion mutants does not directly query the variation, but it queries the outcomes of the variation. For example, $RPI1^{S288c}$ allows S288c^{FLO8} to bypass the fMAPK pathway whereas $RPI1^{Sigma}$ cannot do the same for Sigma. This results in the entire fMAPK pathway showing a strain specific effect. The difference in fMAPK utilization serves to illustrate how one variant might lead to multiple genes possessing a strain specific effect. The ability to detect changes in utilization of whole pathways (e.g. fMAPK) or complexes (e.g. Swr1 complex, see Appendix A1) is a significant benefit because it provides an understanding of the molecular mechanism for causal variants.

At the same time, the fact that the variants are not themselves being queried makes it unlikely that the causal variant will be immediately evident. In support of this hypothesis, to discover *RPI1*'s role in agar adhesion required the design of an additional selection based on the information gained from the deletion screens. While the selection was successful, there were two assumptions in the design of the selection that could have caused it to fail. First, one gene needed to be sufficient to bypass the fMAPK pathway and second, the S288c allele needed to be dominant. These assumptions are not required in traditional linkage mapping and future studies of natural variation will benefit if they can leverage the complimentary benefits of the deletion libraries and linkage mapping.

Natural variation in agar adhesion

Much of our understanding of biology comes from studies in relatively few strain backgrounds, and often only one strain background is used for studies in a particular field. The study of yeast adhesion is a no exception as it has been exclusively studied in the Sigma background in which it was originally discovered (GIMENO *et al.* 1992). This study raises the question of how much of our understanding of yeast adhesion and biology as a whole is biased by strain usage.

The role of the fMAPK pathway in Sigma agar adhesion and *FLO11* expression has been well characterized (CHEN and THORNER 2010; LIU *et al.* 1993; ROBERTS and FINK 1994) and how the activation of the fMAPK is insulated from activation of the pheromone response MAPK pathway has been influential to our understanding of signaling pathway specificity (CHEN and THORNER 2007; MADHANI and FINK 1997; MADHANI *et al.* 1997; SCHWARTZ and MADHANI 2004; SCHWARTZ and MADHANI 2006; WESTFALL and THORNER 2006). Yet the principles that

have been uncovered may not hold true for the S288c where the fMAPK pathway is dispensable for agar adhesion.

If agar adhesion had been studied in an S288c^{FLO8} background, then the role for the fMAPK pathway in the regulation of *FLO11* may not have been discovered. Similarly, until this study, the role and *RPI1* in *FLO11* regulation had not been found, despite several screens in Sigma that examined agar adhesion (GIMENO and FINK 1994; JIN *et al.* 2008; SUZUKI *et al.* 2003; VOYNOV *et al.* 2006).

The discovery of over 500 genes that have strain specific effects in agar adhesion shows that the strain specificity seen with the fMAPK pathway is not an isolated phenomenon but is likely to be generally applicable. Finding more of the genes responsible for the differences in agar adhesion between S288c^{FLO8} and Sigma could lead to an understanding of basic differences between the two strains. For example, the Swr1 chromatin remodeling complex has an S288c^{FLO8} specific adhesion defect (Appendix A1). The adhesion defect is not due to altered *FLO11* expression and therefore must affect gene expression of a process downstream of *FLO11*'s role in agar adhesion; however this change is specific to S288c^{FLO8}. Given the role for the Swr1 complex in modifying chromatin, this finding suggests that the chromatin states in S288c and Sigma differ. Performing transcriptional profiling on *htz1*Δ strains in S288c and Sigma, along with ChIP-seq to compare Htz1p distributions in S288c and Sigma could shed light on how chromatin structure differs between these two strains.

While *RPI1*^{S288c} can bypass the fMAPK pathway, crosses between S288c^{FLO8} *tec1*Δ and Sigma *tec1*Δ showed that ≥3 genes play a role in fMAPK bypass. Also, the effect of the *RPI1*^{S288c} allele is most pronounced on synthetic media and it has little effect on the rich media that is usually used for adhesion assays. The specificity for synthetic media is likely linked to

the requirement for this media in the selection that pulled out *RPI1*. The cross data and the

media specificity suggest that more fMAPK bypassers have yet to be found, but how can they be

uncovered?

New tools for future linkage studies

As previously mentioned, coupling deletion library screens and linkage studies in yeast

could significantly increase our understanding of biology beyond what either could do on its

own. In addition, linkage studies will benefit from the recent development of X-QTL analysis

that was described in Chapter 1 (EHRENREICH *et al.* 2010). In principle, X-QTL should make

mapping any selectable trait as simple as doing two microarrays or deep sequencing runs.

Unfortunately, X-QTL requires an initial selection to enrich for haploid progeny of the same

mating type and this selection has limited X-QTL analyses to the BY x RM cross (EHRENREICH

et al. 2010; TORABI and KRUGLYAK 2011). This selection is done using a set of markers

originally developed for synthetic gene array (SGA) analysis in yeast (TONG and BOONE 2006).

The first marker is the *HIS3* gene under the control of the *MAT*a specific promoter of *STE2* and

allows for the selection of *MAT*a haploids based on His prototrophy. This construct is used to

disrupt the recessive canavanine resistance marker *CAN1*. Lastly, the recessive thialysine

resistance gene *LYP1* is also deleted. The combination of *can1*Δ and *lyp1*Δ helps to select

against unsporulated diploid cells.

While these three markers enrich for *MAT*a haploids, there are three major drawbacks

that have limited the utility of X-QTL analysis. First, the resulting *MAT*a progeny have three

markers that could have pleiotropic effects. Second, using the SGA markers fixes three loci in

the resulting pool – MAT, *CAN1*, and *LYP1*. Any moderate-affect variants near these loci may

go undetected because of the strong selection for these loci. Third, the markers require a growth

selection step and leads to enrichment of genotypes that generally grow better regardless of the

trait being examined.

To address these problems, I developed the FASTER MT method to enrich for *MAT*α

haploids (Appendix A-2). This method results in a pool of markerless progeny where only one

locus is selected for and it has is no growth requirement. This method will provide a significant

benefit to future X-QTL experiments and it allows the use of X-QTL not only RM11 but also

Sigma and any other strain background. Combining X-QTL in Sigma with the deletion library

will allow future studies to rapidly and comprehensively access the natural variation in these

strains.

For examining *FLO11* expression, the *HIS3-PEST* construct was key to finding *RPII*^{S288c},

and it should be possible to utilize *HIS3-PEST* the construct for X-QTL analysis to map

additional loci that participate in fMAPK bypass. However, the requirement that the His$^+$

selection be performed on synthetic media is a major disadvantage. To address this problem I

developed constructs where, instead of *HIS3-PEST*, different drug resistance genes were under

the control of the *FLO11* promoter. Initially none of the constructs were suitable because they

required high concentrations of the drug to produce a growth difference between Sigma wild-

type and *tec1*Δ strain. This is similar to what was seen if the *HIS3* gene is used without the

PEST modificiation. Therefore, I appended the *CLN2* PEST sequence onto the different drug

resistance genes and found the difference between a Sigma wild-type and *tec1*Δ is easily

visualized using the NAT-PEST construct which confers resistance to the drug nourseothricin

(Figure 4-1A). When this construct is used in a cross between S288c^{FLO8} *tec1*Δ and Sigma

*tec1*Δ, the Nat$^+$ phenotype is rare and highly resistant segregants can be selected for on higher

concentrations of nourseothricin (Figure 4-2B). Since NAT-PEST construct can be selected for

on rich media, it will be more suitable than the *HIS3-PEST* construct for locating additional

fMAPK bypassers by transformation or X-QTL analysis. With the combination of FASTER and

the NAT-PEST, future studies will be able to build on the deletion library screens and dissect out

the adhesion differences to an extent that was not previously possible.

While both the *HIS3-PEST* and *NAT-PEST* constructs were developed to discover

bypassers of the fMAPK pathway, they have the potential to be generally useful selectable

constructs. For *FLO11* regulation, they can be used to uncover more of the causal variants

associated with the 500 genes with strain specific adhesion defects. Beyond *FLO11* regulation,

they are potentially useful tools for understanding the transcriptional regulation of any promoter.

Many genes differ in intragenic tandem repeat lengths

Both the *RPI1* and *MIT1* genes obtained from the *HIS3-PEST* selection have intragenic

tandem repeats, and the size of the repeat region differs between the S288c and Sigma alleles.

Previous studies have shown that cell surface genes in yeast often have large intragenic tandem

repeats, and changing the size of these repeats has phenotypic consequences (VERSTREPEN *et al.*

2005). The repeat units in *RPI1* and *MIT* are much smaller than was examined in Verstrepen et

al. (2005) and we asked if they had missed genes by applying highly stringent conditions in their

bioinformatics screens.

The difficulty in finding repeats lies in deciding what is and is not a repeat. Factors

include the number of bases repeated, how many times the unit is repeated, and how much

homology the repeats units must have with each other. To circumvent these issues we leveraged

the high quality genome sequence of Sigma and S288c. We allowed the genomes to tell us what

to look for by first asking how many ORFs differed in size between S288c and Sigma, and then asking how many of those genes differed because of a change in the length of intragenic repeats. 107 genes differed in size by multiples of 3 bases, and visual inspection of nucleotide alignments and experimental validation by PCR suggested that the majority differed by changes in repeat size. The genes with repeat length changes are spread across a wide variety of functions and could impact many different processes.

For a number of cell surface genes changing the size of intragenic tandem repeats has phenotypic consequences (FIDALGO *et al.* 2006; FIDALGO *et al.* 2008; SHEETS and ST GEME 2011; VERSTREPEN *et al.* 2005). If there are similar phenotypic consequences for the 107 genes that we have identified, then among the progeny of a cross between S288c and Sigma there is the potential for 2^{107} different combinations of these length polymorphisms. This level of variation in single cross is larger than the estimated number of number of bacterial cells on the planet (5×10^{30}) (WHITMAN *et al.* 1998).

Figure 5-1 | **NAT-PEST drug selectable marker for gene expression.** All strains shown were homozygous for the *FLO11pr*-NAT-PEST construct. (A) Eighteen tetrads from a Sigma diploid heterozygous for *tec1Δ* shows that the construct can effectively distinguish *tec1Δ*::KanMX from *TEC1* strains. All strains that are Nat- are also *tec1Δ*. Two examples are circled. (B) A hybrid S288c^{FLO8} / Sigma *tec1Δ/tec1Δ* strain shows a heterogeneous segregation of Nat+ that is consistent with the influence of natural variation on *FLO11* expression. Higher expression can be more selected for by increasing the concentration of clonNAT. One highly resistant segregant circled.

A.

Sigma *TEC/tec1::KANMX4, FLO11pr-NATMXPEST/FLO11pr-NATMXPEST*
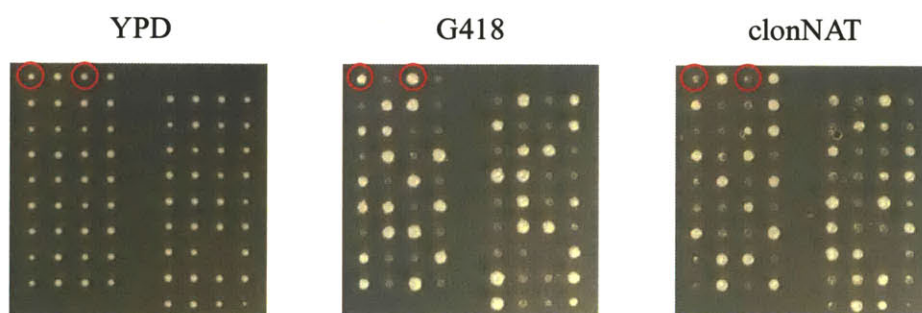
B.

Sigma/S288c^{FLO8} *tec1::KANMX4/tec1::KANMX4, FLO11pr-NATMXPEST/FLO11pr-NATMXPEST*
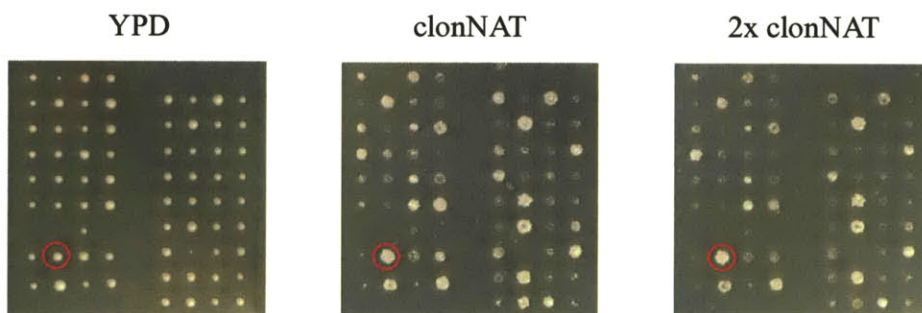
References

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

CHEN, R. E., and J. THORNER, 2007 Function and regulation in MAPK signaling pathways: lessons learned from the yeast Saccharomyces cerevisiae. Biochim Biophys Acta **1773**: 1311-1340.

CHEN, R. E., and J. THORNER, 2010 Systematic epistasis analysis of the contributions of protein kinase A- and mitogen-activated protein kinase-dependent signaling to nutrient limitation-evoked responses in the yeast Saccharomyces cerevisiae. Genetics **185**: 855-870.

DEMOGINES, A., E. SMITH, L. KRUGLYAK and E. ALANI, 2008 Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. PLoS Genet **4**: e1000123.

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

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

EHRENREICH, I. M., N. TORABI, Y. JIA, J. KENT, S. MARTIS *et al.*, 2010 Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature **464**: 1039-1042.

FIDALGO, M., R. R. BARRALES, J. I. IBEAS and J. JIMENEZ, 2006 Adaptive evolution by mutations in the FLO11 gene. Proc Natl Acad Sci U S A **103**: 11228-11233.

FIDALGO, M., R. R. BARRALES and J. JIMENEZ, 2008 Coding repeat instability in the FLO11 gene of Saccharomyces yeasts. Yeast **25**: 879-889.

GIMENO, C. J., and G. R. FINK, 1994 Induction of pseudohyphal growth by overexpression of PHD1, a Saccharomyces cerevisiae gene related to transcriptional regulators of fungal development. Mol Cell Biol **14**: 2100-2112.

GIMENO, C. J., P. O. LJUNGDAHL, C. A. STYLES and G. R. FINK, 1992 Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. Cell **68**: 1077-1090.

JIN, R., C. J. DOBRY, P. J. MCCOWN and A. KUMAR, 2008 Large-scale analysis of yeast filamentous growth by systematic gene disruption and overexpression. Mol Biol Cell **19**: 284-296.

KELLIS, M., N. PATTERSON, M. ENDRIZZI, B. BIRREN and E. S. LANDER, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423**: 241-254.

LEVDANSKY, E., J. ROMANO, Y. SHADKCHAN, H. SHARON, K. J. VERSTREPEN *et al.*, 2007 Coding tandem repeats generate diversity in Aspergillus fumigatus genes. Eukaryot Cell **6**: 1380-1391.

LITI, G., D. M. CARTER, A. M. MOSES, J. WARRINGER, L. PARTS *et al.*, 2009 Population genomics of domestic and wild yeasts. Nature **458**: 337-341.

LIU, H., C. A. STYLES and G. R. FINK, 1993 Elements of the yeast pheromone response pathway required for filamentous growth of diploids. Science **262**: 1741-1744.

MADHANI, H. D., and G. R. FINK, 1997 Combinatorial control required for the specificity of yeast MAPK signaling. Science **275**: 1314-1317.

MADHANI, H. D., C. A. STYLES and G. R. FINK, 1997 MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation. Cell **91**: 673-684.

NOGAMI, S., Y. OHYA and G. YVERT, 2007 Genetic complexity and quantitative trait loci mapping of yeast morphological traits. PLoS Genet **3**: e31.

ROBERTS, R. L., and G. R. FINK, 1994 Elements of a single MAP kinase cascade in Saccharomyces cerevisiae mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev **8**: 2974-2985.

SCHACHERER, J., J. A. SHAPIRO, D. M. RUDERFER and L. KRUGLYAK, 2009 Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. Nature **458**: 342-345.

SCHWARTZ, M. A., and H. D. MADHANI, 2004 Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. Annu Rev Genet **38**: 725-748.

SCHWARTZ, M. A., and H. D. MADHANI, 2006 Control of MAPK signaling specificity by a conserved residue in the MEK-binding domain of the yeast scaffold protein Ste5. Curr Genet **49**: 351-363.

SHEETS, A. J., and J. W. ST GEME, 3RD, 2011 Adhesive activity of the haemophilus cryptic genospecies cha autotransporter is modulated by variation in tandem Peptide repeats. J Bacteriol **193**: 329-339.

SINHA, H., L. DAVID, R. C. PASCON, S. CLAUDER-MUNSTER, S. KRISHNAKUMAR *et al.*, 2008 Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. Genetics **180**: 1661-1670.

STEINMETZ, L. M., H. SINHA, D. R. RICHARDS, J. I. SPIEGELMAN, P. J. OEFNER *et al.*, 2002 Dissecting the architecture of a quantitative trait locus in yeast. Nature **416**: 326-330.

SUZUKI, C., Y. HORI and Y. KASHIWAGI, 2003 Screening and characterization of transposon-insertion mutants in a pseudohyphal strain of Saccharomyces cerevisiae. Yeast **20**: 407-415.

TAN, J. C., A. TAN, L. CHECKLEY, C. M. HONSA and M. T. FERDIG, 2010 Variable numbers of tandem repeats in Plasmodium falciparum genes. J Mol Evol **71**: 268-278.

TONG, A. H., and C. BOONE, 2006 Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol Biol **313**: 171-192.

TORABI, N., and L. KRUGLYAK, 2011 Variants in SUP45 and TRM10 underlie natural variation in translation termination efficiency in Saccharomyces cerevisiae. PLoS Genet **7**: e1002211.

VERSTREPEN, K. J., A. JANSEN, F. LEWITTER and G. R. FINK, 2005 Intragenic tandem repeats generate functional variability. Nat Genet **37**: 986-990.

VOYNOV, V., K. J. VERSTREPEN, A. JANSEN, V. M. RUNNER, S. BURATOWSKI *et al.*, 2006 Genes with internal repeats require the THO complex for transcription. Proc Natl Acad Sci U S A **103**: 14423-14428.

WESTFALL, P. J., and J. THORNER, 2006 Analysis of mitogen-activated protein kinase signaling specificity in response to hyperosmotic stress: use of an analog-sensitive HOG1 allele. Eukaryot Cell **5**: 1215-1228.

WHITMAN, W. B., D. C. COLEMAN and W. J. WIEBE, 1998 Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A **95**: 6578-6583.

WINZELER, E. A., H. LIANG, D. D. SHOEMAKER and R. W. DAVIS, 2000 Functional analysis of the yeast genome by precise deletion and parallel phenotypic characterization. Novartis Found Symp **229**: 105-109; discussion 109-111.

Appendix A1

Deletion library screens in two strain backgrounds reveals novel regulation of adhesion phenotypes.

Abstract

The properties of a microbes' cell surface play a crucial role in their survival and their ability to cause diseases. Changes in the expression of cell surface proteins can allow for microbes to adhere to plastics, form biofilms and evade the immune system, but how microbes regulate the repertoire of proteins on their surface is not well understood. In the budding yeast *Saccharomyces cerevisiae*, expression of the *FLO* family of cell surface proteins can lead to a variety of different cell surface properties. For example, expression of the *FLO1* gene will lead to cells forming tight aggregates, while expression of *FLO11* allows haploid cell to adhere to substrates. While many studies have examined *FLO11* regulation and found that a complex network of signaling pathways controls its expression, little is known about how cells regulate different *FLO* genes in response to different situations. There is also a large amount of heterogeneity in what *FLO* genes are present in different strains, and it is possible that different strains will differ in the regulation the *FLO* genes. To understand how *S. cerevisiae* controls *FLO* gene expression, we constructed an S288c deletion library that was competent for *FLO* gene expression. Using this S288c deletion library we found that perturbations of the cell wall affect *FLO11* expression but not *FLO1* expression. We also compared all of the functions required for agar adhesion in S288c to all the function for agar adhesion in another genetic background, Σ1278b (Sigma) using deletion libraries for each strain. The results from the two deletion libraries showed that the majority of agar adhesion regulators were strain specific. Further analysis showed that the regulation of agar adhesion is complicated by the strain specific requirement of the Swr1 complex in S288c and the filamentation mitogen activated kinase pathway in Sigma. These comprehensive sets of screens illustrate the complexity of cell surface regulation in *S. cerevisiae*.

Introduction

The *Saccharomyces cerevisiae FLO* gene family is made up cell-surface glycoproteins and can confer a range of cell-cell and cell-substrate adhesion properties to yeast cells (GUO *et al.* 2000), and both processes have important implications for utility and virulence of *S. cerevisiae*. Cell-cell adhesion in yeast is called 'flocculation' and is frequently used in the brewing industry where it is desirable to have yeast form macroscopic aggregates ('flocs') and sediment at the end of a fermentation (VERSTREPEN *et al.* 2003). Cell-substrate adhesion is often

associated with morphological changes in the yeast (KRON *et al.* 1994; ROBERTS and FINK 1994), and is an important element of the virulence of pathogenic yeast (LO *et al.* 1997).

The two processes of flocculation and cell-substrate adhesion are often mediated by different members of the *FLO* gene family. The *Saccharomyces cerevisiae* strain S288c has five genes that belong to this family: *FLO1*, *FLO5*, *FLO9*, *FLO10*, and *FLO11*. The genes *FLO1*, *FLO5*, and *FLO9* are all involved in flocculation, but have little effect on cell-substrate adhesion (GUO *et al.* 2000). Conversely, *FLO11* has a prominent role in cell-substrate adhesion and very little ability to induce flocculation. *FLO10* is the only *FLO* gene that can participate in both processes.

Yeast adherence to substrates is often measured on the agar plates on which yeast are routinely grown and is termed invasive growth or agar adhesion (ROBERTS and FINK 1994). Agar adhesion in *S. cerevisiae* is dependent on many different processes. One of the key genes involved in agar adhesion is the small GTP binding protein Ras2p (GIMENO *et al.* 1992). Ras2p activates the filamentation mitogen activated protein kinase (fMAPK) pathway and the protein kinase A (PKA) pathway (AHN *et al.* 1999; GIMENO *et al.* 1992; MADHANI *et al.* 1999; MOSCH *et al.* 1999; PAN *et al.* 2000; PAN and HEITMAN 1999; ROBERTS *et al.* 1997).

S. cerevisiae has three different PKAs, Tpk1p, Tpk2p, and Tpk3p. Tpk1p does not have an effect on yeast adhesion, but Tpk2p is required for agar adhesion and Tpk3p is inhibitory toward agar adhesion. The activation of Tpk2p leads to activation of the transcription factor Flo8p and inactivation of the repressor Sfl1p (FURUKAWA *et al.* 2009; PAN and HEITMAN 1999).

The fMAPK pathway shares many components with the pheromone response and hyperosmotic glycerol MAPK pathways. In particular, Ste11p is a common component of all three pathways, and Ste7p and Ste12p are shared between the pheromone response and

filamentation MAP kinase pathways (LIU *et al.* 1993). In the fMAPK pathway, activation of the pathway eventually leads to activation of the MAPK Kss1p which phosphorylates the transcription factors Ste12p and Tec1p.

The transcription factors for both the PKA and the fMAPK pathways converge on the large promoter of *FLO11* and have additive effects on *FLO11* transcription (CHEN and THORNER 2010; RUPP *et al.* 1999). However, the regulation of agar adhesion is not limited to these two pathways. The transcriptional activator, Phd1p promotes agar adhesion in a pathway that involves Sok2p and the kinase Yak1p (MALCHER *et al.* 2011). The repressors Nrg1p and Nrg2p also regulate agar adhesion in a pathway that involves the Snf1p protein kinase (KUCHIN *et al.* 2002; VYAS *et al.* 2003). In addition the transcription factors *MSS11*, *MSN1*, and *ASH1* also play a role in agar adhesion and *FLO11* transcription, but their roles have yet to be fully elucidated (CHANDARLAPATY and ERREDE 1998; PAN *et al.* 2000; VAN DYK *et al.* 2005).

While many signaling pathways regulate *FLO11* expression, their ability to activate *FLO11* transcription is dependent upon the chromatin state at the *FLO11* promoter. The epigenetic silencing of *FLO11* was first shown to be mediated by the histone deacetylase Hda1p and the repressor Sfl1p (HALME *et al.* 2004). This epigenetic silencing results in a genetically homogenous population that varies in the expression of Flo11p. Subsequent studies have uncovered a pair of intergenic noncoding RNAs within the *FLO11* promoter that also regulate *FLO11*'s variegated expression (BUMGARNER *et al.* 2009). These noncoding RNA are under the control of the Rpd3L histone deacetylase complex. Finding that the Rpd3L complex silenced expression of an interfering RNA in the *FLO11* promoter helped to explain how a histone deacetylase could promote the transcription of *FLO11*.

In addition to factors that influence transcription initiation, Flo11p expression is also regulated at the stages of transcription elongation, RNA localization, and protein translation (VOYNOV *et al.* 2006; WOLF *et al.* 2010). Voynov et al. (2006) showed that the THO complex is required for the transcription elongation of *FLO11*, and this requirement is due to a repeated sequence within the *FLO11* ORF. After *FLO11* is transcribed, the translational repressor Khd1p can regulate agar adhesion by suppressing Flo11p translation in a process that also requires the repeat sequences in the *FLO11* ORF (WOLF *et al.* 2010).

All of the previously mentioned pathways have been shown to influence agar adhesion through altering Flo11p expression. In the Sigma strains used to study agar adhesion *FLO11* is the only *FLO* gene expressed, and is required for agar adhesion. This has led to Flo11p being the best studied yeast adhesion molecule. Flo11p is often thought of as the structural protein necessary for agar adhesion, yet how Flo11p mediates agar adhesion is not clear (LO and DRANGINIS 1998; VERSTREPEN and KLIS 2006).

While the molecular control of agar adhesion has been extensively studied, the control over flocculation is far less well understood. Like agar adhesion, flocculation requires the transcription factors *FLO8* and *MSS11* for expression (BESTER *et al.* 2006; KOBAYASHI *et al.* 1999). *HDA1*, the COMPASS histone methylation complex, and the Swi-Snf chromatin remodeling complex have also been shown to be involved in regulating flocculation (DIETVORST and BRANDT 2008).

The molecular mechanism behind flocculation is thought to be relatively simple. A flocculating cell expresses one of the *FLO* genes (e.g. *FLO1*) that can form lectin-bonds with the mannan in the yeast cell wall (VERSTREPEN and KLIS 2006). This process is dependent upon the presence of calcium ions and can be partially inhibited by different sugars (STRATFORD 1989;

STRATFORD and ASSINDER 1991). Chelating the calcium in the media will cause the flocs to disperse. The loss of flocculation can then be reversed by supplementing with additional calcium. The reversibility of flocculation makes it distinct from cell aggregates that arise from incomplete separation of the mother and daughter cells after cytokinesis. While the physical properties of flocs have been extensively studied, studies examining the genes involved in flocculation regulation are complicated by the fact that most laboratory strains do not flocculate, and therefore the studies are often done in strains with uncharacterized genomes.

The Σ1278b (Sigma) strain is used in the majority of the studies examining agar adhesion, and the only *FLO* gene expressed in this strain is *FLO11* (GUO *et al.* 2000). Sigma is considered more like wild yeast than most laboratory strains, yet it does not flocculate.

S288c, despite being a well characterized strain of *S. cerevisiae*, and the references strain for the *S. cerevisiae* genome. has not been used to study adhesion or flocculation because it contains a nonsense mutation in the transcription factor *FLO8* (LIU *et al.* 1996). In an S288c *FLO8* strain, both *FLO11* and *FLO1* are expressed, and the strain will adhere to agar and flocculate (Figure A1-1) (LIU *et al.* 1996).

To obtain a comprehensive, genome-wide understanding of the factors regulating yeast adhesion properties, I performed two screens to examine the two different aspects of yeast adhesion. The first screen looked for genes that regulate flocculation in a *FLO8* version of S288c, and these results were compared to the second screen where agar adhesion was examined in this same strain background. In addition, my data sets were compared to an adhesion screen done on the Sigma deletion library. These two comparisons have showed that the regulation of yeast adhesion is more complex than previous studies have suggested.

Results

Deletion library screens in two strain backgrounds reveal novel regulation of adhesion phenotypes

The standard S288c deletion library cannot be examined for adhesion phenotypes because the S288c progenitor does not express any member of the *FLO* gene family due to a nonsense mutation in the *FLO8* gene (LIU *et al.* 1996). Therefore, in order to utilize the S288c deletion library to identify functions required for yeast adhesion I transformed the 4705 nonessential deletion strains in the standard S288c *flo8* library with a CEN/ARS plasmid carrying the Sigma *FLO8* gene under the control of its own promoter. The 4633 FLO8+ deletion strains successfully recovered from these transformations formed the S288c *FLO8+* (S288c^{FLO8}) deletion library.

Visual screens for increased and decreased agar adhesion are well established, but no protocols existed to perform a genome-wide analysis of flocculation (Figure A1-2A). While a loss of flocculation could be determined visually, increased flocculation cannot be visually detected because wild-type S288c^{FLO8} flocculation is too strong. Therefore, I developed an assay that relied upon the difference in sedimentation rates of cells resuspended in buffers with or without calcium. This assay allowed for a quantitative assessment of flocculation and could be performed in 96-well plates with modest throughput. One crucial addition to the assay was the use of mannose to partially inhibit flocculation. This increased the dynamic range of the assay and allowed me to assay for both decreased and increased flocculation (Figure A1-2B).

I screened the S288c^{FLO8} deletion library for alterations in adhesion and flocculation. These screens revealed 664 different deletions that alter yeast adhesive properties: 316 deletions cause decreased agar adhesion (Ahs-), 88 deletions cause increased agar adhesion, 185 deletions

cause decreased flocculation, and 200 deletion cause increased flocculation (Hfl$^+$) (Figure A1-2C).

I also compared the Ahs$^-$ deletions in S288c^{FLO8} to the Ahs$^-$ deletions in Sigma. This comparison resulted in an additional 283 deletions that affect agar adhesion, but only in Sigma. Gene ontology term enrichment was used to see if certain biological processes preferentially affect certain phenotypes (Table A1-1 and Table 3-1).

Ahs$^-$ deletions in S288c^{FLO8} are enriched for chromatin modifiers. Among this set of chromatin modifiers are members of the Rpd3L complex, namely *CTI6, SIN3, DEP1, ASH1* and *SDS3*. This complex has been recently shown to regulate adhesion in Sigma through a complex mechanism involving non-coding RNAs in the *FLO11* promoter (BUMGARNER *et al.* 2009). This mechanism appears to be common between the two strains. Additionally, the S288c^{FLO8} Ahs-screen identified *SWC5, SWR1, VPS71, VPS72, YAF9* and *HTZ1*, all members of the Swr1p complex that loads the histone H2A variant *HTZ1* onto chromatin (KOBOR *et al.* 2004; KROGAN *et al.* 2003; MIZUGUCHI *et al.* 2004). Also enriched among the S288c^{FLO8} Ahs$^-$ are genes known to affect filamentatous growth. This set includes many of the well described regulators of *FLO11* expression, such as *TPK2, MSS11, ASH1, PHD1* and the THO complex member *THP2*, but this set is missing several key members. In particular, members of the filamentation MAPK (fMAPK) pathway are not present in the set of S288c^{FLO8} Ahs$^-$ deletions.

The S288c^{FLO8} screen for increased agar adhesion identified several known repressors of agar adhesion. The set of increased agar adhesion mutants is enriched for genes that are involved with bud site selection, including *BUD3, RSR1, ERV14, BUD2,* and *AXL2*. This result is consistent with previously published reports of mutations that increase adhesion in Sigma (PALECEK *et al.* 2000). Knockouts of *BUD3* and *AXL1* increase yeast agar adhesion in the Sigma

strain. Furthermore, both my screen and Palecek et al. (2000) saw increased adhesion in mutants of *SIR2*, *SIR4*.

The set of deletions resulting in non-flocculant S288c^{FLO8} is enriched for genes involved in response to heat and chromatin silencing. The chromatin silencing genes are all components of the Rpd3L complex, the same complex that promotes adhesion. The fact that both flocculation and adhesion rely upon the Rpd3L complex suggests that *FLO1* may also be regulated by non-coding RNAs in its promoter.

No GO terms are enriched among the set of Hfl$^+$ deletions, but visual inspection revealed genes involved in glycosylation or cell wall biogenesis (Table A1-2) negatively regulate flocculation. Many of these genes also have an Ahs$^-$ phenotype, suggesting they have contrasting effects on *FLO1* and *FLO11* regulation.

A comparison of S288c$_{FLO8}$ and Sigma Ahs$^-$ deletions yielded the unanticipated result that a minority of deletions were Ahs$^-$ in both strains. There are 270 deletions that are Ahs$^-$ specifically in S288c^{FLO8} and 283 deletions that are Ahs$^-$ only in Sigma, but only 46 deletions are Ahs$^-$ in both. The 46 common Ahs$^-$ deletions are enriched for many of the known regulators of *FLO11* expression.

To eliminate the possibility that the differences in regulation between S288c^{FLO8} and Sigma are due to using a plasmid based *FLO8* in S288c, I constructed S288c^{FLO8} strains where *FLO8+* replaces the endogenous *flo8-* allele. This strain expresses *FLO8* to the same level as Sigma and all of my subsequent studies utilized these S288c^{FLO8} strains.

gas1Δ affects adhesion on multiple levels

135

Many of the deletions in S288c^{FLO8} that result in an Hfl$^+$ phenotype are in genes that maintain the cell wall (Table A1-2). In addition, deletions of *ALG3*, *CHS3*, *GAS1*, *LAS21*, *MNN11*, and *WSC1* also have an Ahs$^-$ phenotype, which suggests they have opposing effects on *FLO1* and *FLO11* expression. In further support of this hypothesis, deletion of *GAS1* has been previously implicated in *FLO11* regulation (Voynov, unpublished results).

Gas1 is a β-1,3-glucanosyltransferase that is important for the maintenance of the major cell wall polysaccharide β -1,3 glucan (MOUYNA *et al.* 2000). Null mutants of *GAS1* have a weakened cell wall and an increased sensitivity to cell wall affecting drugs (DAGUE *et al.* 2010; RAM *et al.* 1998; RAM *et al.* 1994). The *GAS1* deletion was identified from my screens with an Ahs$^-$ and Hfl$^+$ phenotype in S288c^{FLO8}, but the same deletion showed no phenotype in the Sigma screen.

To examine the possibility that perturbations of the cell wall have contrasting effects on *FLO1* and *FLO11* I further examined the effects of *GAS1* on these genes. I used qPCR to examined the levels of *FLO1* and *FLO11* transcript in wild-type and *gas1Δ* strains. The increase in flocculation occurs without a corresponding increase in *FLO1* transcript (Figure A1-4B), which suggests that changes in the physical properties of the cell surface are responsible for the increased flocculation, rather than an altered of regulation of *FLO1*. However, unlike *FLO1*, *FLO11* levels are significantly lowered in this strain, suggesting that the *gas1Δ* specifically affects the regulation of *FLO11* and not *FLO1* (Figure A1-4A). The *gas1Δ* associated decrease in *FLO11* transcript is present in both S288c^{FLO8} and Sigma, despite Sigma's higher level of adhesion.

To examine if the decrease in *FLO11* transcript is responsible for the loss of adhesion in the *gas1Δ* strain, a *TEFpr-FLO11* construct was placed into an S288c *gas1Δ* strain. This

construct eliminates the transcriptional control of *FLO11* by constitutively expressing it from the

TEF promoter. Despite the constitutive transcription of *FLO11* from this construct, an S288c

TEFpr-FLO11, *gas1Δ* strain is non-adherent (Figure A1-4C). These data show that the *gas1Δ*

mutation affects adhesion by two mechanisms, altering *FLO11* transcript levels, and an

additional mechanism that is independent of *FLO11* transcription.

Two possible mechanisms can account for the decrease in *FLO11* transcript in a *gas1Δ*:

FLO11 transcript could decrease due to a general defect in the transcription of *FLO11* or,

alternatively, an increase in the epigenetic silencing of the *FLO11* promoter could lead to fewer

cells expressing *FLO11*. These two possible mechanisms were examined using two different

transcriptional fusions to the *FLO11* promoter: *FLO11pr-URA3*, and *FLO11pr-GFP*.

The *FLO11pr-URA3* construct is useful because there are positive and negative selections

against Ura3p enzymatic activity. Cells expressing Ura3p can be positively selected for by

plating onto media lacking uracil (-Ura), where only cells producing the Ura3p protein will grow.

Cells expressing Ura3p can be selected against by plating onto media containing 5-fluoroorotic

acid (5-FOA), where only cells without Ura3p enzymatic activity can grow. An isogenic

population of yeast with the *FLO11pr-URA3* construct is able to grow on both -URA and 5-FOA

media because variegation of *FLO11* expression results in a mixed population where some cells

are actively transcribing from the *FLO11* promoter and others are not (Figure A1-5A). This

phenomenon is true for both S288c^{FLO8} and Sigma.

Changes in the epigenetic silencing of the *FLO11* promoter can be seen as a change in the

proportion of cells able to grow on -URA versus 5-FOA media. Deletion of *FLO11* regulators,

such as *TEC1* and *STE12*, that do not affect silencing of the *FLO11* promoter but instead affect

transcription levels of an already activated promoter, do not change the proportion of cells that

grow on -URA or 5-FOA plate (HALME *et al.* 2004; OCTAVIO *et al.* 2009). Conversely, genes such as *CTI6* and *HDA1* affect the silencing of the *FLO11* promoter and the deletion of these genes changes the proportion of cells that can grow on -URA or 5-FOA (BUMGARNER *et al.* 2009; HALME *et al.* 2004). When examined by dilution series, a fewer number of S288c[FLO8] *gas1Δ*, *FLO11pr-URA3* cells grown on -URA plates and more cells grow on 5-FOA plates when compared to wild-type (Figure A1-5A). This shift in the population suggests that *gas1Δ* has increased silencing of the *FLO11* promoter.

The increase in *FLO11* promoter silencing in a *gas1Δ* strain is further supported by fluorescent microscopy of *FLO11pr-GFP* expressing strains. In strains with the *FLO11pr-GFP* construct, the epigenetic silencing can be visualized by examining the cells for GFP fluorescence. In wild-type strains, some of the cells will express GFP and others will not (Figure A1-5B) (HALME *et al.* 2004). This variegated expression of GFP is present in both S288c[FLO8] and Sigma strains. In *gas1Δ* mutants of both S288c[FLO8] and Sigma, the number of cells expressing this construct is decreased (Figure A1-5B).

Flow cytometry can be used to obtain quantitative measurements of GFP expressing strains, but *gas1Δ* strains have a defect in mother-daughter separation that results in clumps of cells (POPOLO *et al.* 1993). This separation defect made single cell measurements impossible and the clumps contained a mixture of expressing and non-expressing cells, making analysis of GFP fluorescence in the clump irrelevant.

It is possible to separate the clumps of cells and obtain single cell measurements by digesting away the cell wall with enzymes such as zymolyase or lyticase. Such a treatment should not affect GFP protein stability; however, zymolyase treatment of the wild-type cells with the *FLO11pr-GFP* construct showed that fewer cells were expressing GFP (Figure A1-6). This

is consistent with the hypothesis that cell wall perturbations affect *FLO11* expression. The

decrease in GFP expressing cells upon zymolyase treatment phenocopies the decrease in GFP

expressing cells in the *gas1Δ* strains.

The Swr1 complex specifically affects S288c^{FLO8} adhesion independent of Flo11 expression

The S288c^{FLO8} Ahs⁻ screen identified six members of the Swr1 chromatin remodeling

complex, namely *SWC3*, *SWR1*, *VPS71*, *VPS72*, *YAF9* and *HTZ1*, as positive regulators of

adhesion. The role of the Swr1 complex in adhesion is specific to S288c^{FLO8}. Deletion in any of

these genes leads to an Ahs⁻ phenotype in S288c^{FLO8}, but has no effect on Sigma adhesion. One

exception is the *SWC7* subunit of the complex. *SWC7* shows a Sigma specific adhesion defect

with no defect in S288c^{FLO8}.

I confirmed the adhesion phenotypes of *SWC3*, *SWR1* and *HTZ1* by making clean

deletions of these genes in S288c^{FLO8} and Sigma strains (Figure A1-7). *swc3Δ*, *swr1Δ* and *htz1Δ*

all give a reproducible Ahs- phenotype, but only in S288c^{FLO8}.

The Swr1 complex is a 13 subunit complex that loads the histone H2A variant H2A.Z

(Htz1 in *S. cerevisiae*) onto chromatin (KOBOR *et al.* 2004; KROGAN *et al.* 2003; MIZUGUCHI *et*

al. 2004). Htz1 is present at the promoters of many genes in euchromatin and it is thought to

play a role in preventing gene silencing (MENEGHINI *et al.* 2003; SANTISTEBAN *et al.* 2000;

VENKATASUBRAHMANYAM *et al.* 2007). In *htz1Δ* cells genes near the telomere are repressed and

there is ectopic spread of the Sir2-Sir3-Sir4 silencing complex (MENEGHINI *et al.* 2003). To test

if the Swr1 complex is necessary for *FLO11* expression, I examined *FLO11* transcript levels in

swr1Δ and *htz1Δ* strains. In contrast to the adhesion results, *FLO11* transcript levels remain high

in S288c^{FLO8} *swr1Δ* or S288c^{FLO8} *htz1Δ* (Figure A1-8A and A1-8b). To further reinforce these results, I examined GFP fluorescence in an S288c^{FLO8} *FLO11pr-GFP*, *htz1Δ* strain (Figure A1-8C). There is neither a change in the level of GFP fluorescence per cell, nor the number of cells expressing GFP in the *htz1Δ* strain.

These results suggested that the role of *HTZ1* in adhesion occurs after *FLO11* transcription. To see if S288c^{FLO8} *htz1Δ* strains have a defect in the translation or localization of Flo11p protein, I constructed HA tagged alleles of *FLO11* and visualized Flo11p-HA protein by immunofluorescence. Flo11p-HA protein is expressed in S288c *htz1Δ* strains and properly localizes to the cell surface (Figure A1-9). These data suggest that the defect in adhesion occurs independently of Flo11p expression, in a process that is specific to S288c^{FLO8}.

MAP kinase pathways have strain specific roles in adhesion.

The fMAPK pathway one of the first pathways discovered to regulate the *FLO11* dependent phenotype of filamentation (LIU *et al.* 1993). It was subsequently shown that the fMAPK pathway is required for *FLO11* expression (PAN and HEITMAN 1999; RUPP *et al.* 1999). Members of the fMAPK pathway also participate in other well characterized MAP kinase pathways. In particular, *STE12* participates in the pheromone response pathway and *STE7*, and *STE11* participate in both the pheromone response and the high osmolarity glycerol (HOG) pathway (MADHANI and FINK 1997; MADHANI *et al.* 1997; O'ROURKE and HERSKOWITZ 1998; ROBERTS and FINK 1994). Both the pheromone response and HOG pathways are known to be active in both S288c and in Sigma. Due to their shared components, the fMAPK, pheromone response and HOG pathways have also formed the basis for many studies examining crosstalk and signaling specificity between signaling pathways (CHEN and THORNER 2007; SCHWARTZ and

MADHANI 2004; SCHWARTZ and MADHANI 2006; VINOD and VENKATESH 2007; WESTFALL and THORNER 2006).

The majority of the fMAPK pathway was identified in the Sigma adhesion screen as activators of adhesion, namely *STE7*, *STE11*, *KSS1*, *STE12*, and *TEC1*. It was expected that these same members would also emerge from the S288c^{FLO8} adhesion screen, but they did not. To confirm the adhesion phenotypes, I constructed deletions of fMAPK genes in an S288c^{FLO8} strain with *FLO8* integrated into the genome. I compared the S288c^{FLO8} deletion strains to the same deletions in Sigma. *ste7Δ*, *ste11Δ*, *kss1Δ*, *ste12Δ* and *tec1Δ* are reproducibly Ahs- in Sigma but Ahs+ in S288c^{FLO8} (Figure 3-1A). Consistent with the adhesion phenotype, a Sigma *tec1Δ* has decreased *FLO11* transcript, but an S288c^{FLO8} *tec1Δ* has high levels of *FLO11* (Figure 3-1C).

I hypothesized that in S288c^{FLO8} the other MAPK pathways might take over for the fMAPK for *FLO11* expression. Although none of the other MAP kinases had an S288c^{FLO8} agar adhesion defect in my screens, it is likely that minor defects in adhesion were not detected and the kinases may be partially redundant. Upon more detailed inspection, deletions of the MAPKs for the pheromone response pathway and the HOG pathway, *FUS3* and *HOG1* respectively, have mild adhesion defects that are specific to S288c^{FLO8} (Figure A1-10A). The double mutant has a more severe S288c^{FLO8} adhesion defect, but no defect in Sigma.

The adhesion defects in S288c^{FLO8} *fus3Δ*, *hog1Δ* mutants correlate with a decrease in *FLO11* transcript. S288c^{FLO8} *hog1Δ* strains have a decrease in *FLO11* transcript levels and the S288c^{FLO8} *hog1Δ*, *fus3Δ* double mutant has a further decrease in *FLO11* transcript (Figure A1-10B). The additive effect of the *hog1Δ* and *fus3Δ* mutations on S288c^{FLO8} *FLO11* expression suggests that these two genes are acting in separate pathways. Neither deletion, singly or in

combination, has any effect on Sigma *FLO11* transcript levels, further supporting the hypothesis

that the MAP kinase pathways have been rewired between S288c and Sigma.

Discussion

The systematic deletion screens described in this chapter have identified 947 genes that can potentially affect yeast adhesion. This data set provides a comprehensive view of the regulation of yeast adhesion phenotypes in two different strains. The results from these screens show the presence of many regulatory mechanisms controlling the yeast cell surface. While each screen on its own provides a large amount of data, comparing the data between the screens reveals previously uncharacterized mechanisms by which the *FLO* genes can be regulated. From comparing the genes important in flocculation to those important in adhesion, the screens revealed that deletions altering the cell wall cause an increase in flocculation but a decrease in adhesion. In particular, deletion of *GAS1* has a severe adhesion defect and a loss of *FLO11* transcripts but *FLO1* levels remain at wild-type levels. This illustrates how the two genes can be independently regulated, despite sharing numerous regulatory factors.

It has only recently been shown that *GAS1* has a role in silencing at specific loci (KOCH and PILLUS 2009). Koch and Pillus (2009) found a defect in silencing at the telomeres and increased silencing at the rDNA in a *gas1Δ* and showed a physical interaction between Gas1p and the histone deacetylase Sir2p. These findings suggest a direct role for Gas1p in chromatin silencing. My finding that *gas1Δ* affects the silencing at the *FLO11* promoter is consistent with this study, although the exact mechanism behind Gas1p's role in silencing remains unknown. As many of the factors involved in silencing of *FLO11* have been worked out, *FLO11* may prove to be an ideal locus to study how Gas1p is able to affect chromatin states.

This silencing of the *FLO11* promoter is not the only defect in agar adhesion caused by deletion of *GAS1*. Transcription of *FLO11* in a *gas1Δ* is not sufficient to restore agar adhesion. Only a few studies have examined the post-transcriptional regulation of *FLO11* and it is possible

that in a *gas1Δ* the *FLO11* protein cannot be made (VOYNOV *et al.* 2006; WOLF *et al.* 2010). Alternatively, it is also possible there are yet uncharacterized functions that act downstream of Flo11p protein expression that are required for adhesion. In support of the later possibility, the Swr1 complex is required for adhesion, even in the presence of Flo11p protein.

The Swr1 complex loads the histone variant *htz1* onto chromatin and this variant is associated with preventing gene silencing (MENEGHINI *et al.* 2003). The S288c^{FLO8} adhesion screens pulled out 5 out of 8 viable deletions of members of the Swr1 complex as well as *htz1Δ*. Swr1 complex mutants have an S288c^{FLO8} specific adhesion defect and this strain specific defect suggests that S288c and Sigma have important differences in their chromatin structure.

The characterization of the *gas1Δ* and *htz1Δ* deletions suggests additional roles for chromatin in regulating agar adhesion, but in both cases the agar adhesion defect is most pronounced in S288c^{FLO8}. These are not the only two examples of strain specific regulators of agar adhesion, 553 deletions have strain specific agar adhesion defects. How can there be such large differences in the regulation of agar adhesion when S288c and Sigma have an average of 99.7% homology across the genome? It is possible that there are many small differences in regulation between the two strains, or perhaps a few master regulators have changed and have thus restructured the signaling network for *FLO11*. The result that the entire fMAPK pathway is not needed for S288c^{FLO8} adhesion supports the later hypothesis, but also raises the question; what modifications in the S288c^{FLO8} strain bypass the fMAPK pathway?

Using classical genetic techniques to find the S288c fMAPK bypassers was impractical due to the large amount of natural variation in the regulation of agar adhesion. Instead, using a candidate gene approach I found that two other MAP kinases have an S288c specific adhesion

defect. This suggests that the signaling pathways act differently in S288c^{FLO8} than in Sigma, despite the high degree of homology between these two strains.

It is important to note that many members of the fMAPK pathway are known to be functional in S288c^{FLO8}. Mapping variation relies upon variation present in the genome, but there is no variation in the S288c and Sigma sequences of *TEC1*, *STE12*, *KSS1*, *STE7* or *STE11*. If the impact of natural variation on agar adhesion had been examined using standard or quantitative trait loci mapping techniques, it is very likely that the difference in utilization of the fMAPK pathway would not be immediately evident because both strains have active members of this pathway. Furthermore, the genetic evidence that I have presented in this chapter suggests that wild-type S288c^{FLO8} and Sigma have many loci that can impact agar adhesion. If each locus has an additive effect on agar adhesion, then as more loci affect a trait, the contribution of each locus to that trait decreases. This makes mapping the many loci difficult because the contribution of each locus might be small. However, by examining natural variation through the use of deletion libraries I overcame many of the problems in mapping natural variation. I was able to quickly identify hundreds of mutations with strain specific effects on agar adhesion, and I could easily identify the molecular processes specific to agar adhesion in S288c^{FLO8} and Sigma.

Materials and Methods

Strains, Media, Microbiological Techniques, and Growth Conditions.

Yeast strains used in this study are derived from S288C and Σ1278b. Standard yeast

media were prepared and genetic manipulation techniques were carried out as described

(GUTHRIE and FINK 2002). Standard adhesion assays were carried out by densely patching

strains onto YPD plate. These plates were grown overnight at 30°C and then replica plated onto

YPD plates. The replica plates were grown for 30°C. After three days the plates were washed

by partially filling the petri dish with 10mM EDTA to prevent flocculation followed by gentle

shaking at approximately 75rpm on an orbital shaker. For qPCR and fluorescent microscopy,

cells were grown overnight in liquid YPD media. The next morning the cultures were diluted to

OD_{600} 0.1 in liquid YPD media, and grown to OD_{600} 0.8-1.2 for use in experiments. For

analysis of *FLO11pr-URA3* strains were grown overnight in YPD liquid media, then diluted 1:50

in YPD liquid media and grown to OD_{600} 0.8-1.2. Cell densities were adjusted to OD_{600} 1. Then

cultures were serially diluted 5-fold and plated synthetic complete (SC), SC-Ura, and SC+5-FOA

(0.1%) agar plates. Cultures for GFP measurements were grown overnight in liquid YPD and

then pelleted and resuspended in water. Samples were transferred to a Corning 96 well black

clear-bottom plate and GFP fluorescence was measured in a Tecan Safire2 plate reader. For

zymolyase treatment of *gas1Δ* strains, cells were grown up for fluorescent microscopy then

pelleted and resuspended in 10mM tris with or without Zymolyase 100T at a concentration of

1.5U/ml. Cells were incubated at 30°C for half an hour then imaged.

Yeast strain construction

To create the S288c^{FLO8} deletion collection the S288c haploid MATa deletion collection

(Invitrogen 95401.H2P) was transformed with plasmid pHL1 (LIU *et al.* 1996) using previously

published protocols (VOYNOV *et al.* 2006).

To create the S288c^{FLO8} wild-type strains used in subsequent studies, BY4742

(BRACHMANN *et al.* 1998) was transformed with the *FLO8* containing *URA3* marked integrating

plasmid pBC6 cut with BglII. Correct transformants then had the marker looped out and *FLO8+*

recombinants were screened for flocculation. This produced yBC9 which was crossed to

BY4741 and the resulting diploid was sporulated to give yBC37.

Yeast strains carrying gene deletions were constructed by PCR amplification of

kanamycin-resistance gene cassettes from the yeast deletion library (WINZELER *et al.* 1999) with

approximately 200 bases of flanking sequence.

FLO11pr-URA3 strains were constructed by PCR of the *URA3* ORF with primers

containing 50-mer flanking sequences that are homologous to the 5' and 3' sequences flanking

the target ORF. The *URA3* amplified constructs completely replaced the target ORF with the full

ORF of *URA3* (ATG-Stop), so that *URA3* expression would be controlled by the target ORF

promoter. Correct integration of the *URA3* construct after transformation was verified by PCR.

FLO11::HA alleles were generated as described (Guo et al., 2000).

Adhesion screens

To perform the agar adhesion screens of the S288c^{FLO8} and Sigma deletion collections the

deletion collections were spotted onto solid media (SC-Ura for the S288c^{FLO8} collection and

YPD for the Sigma collection) and then grown up overnight at 30°C. The following day the

strains were transferred in triplicate onto YPD plate using a 96 prong frogger and grown for three days at 30°C. Plates were washed by placing them in Tupperware filled with 10mM EDTA and gently shook. The plates were scored visually for adhesion defects.

To perform the flocculation screen on the S288c^{FLO8} deletion collection, the strains were grown overnight in 96-well format in 200μl of liquid SC-Ura media in. The following day the cells were spun down, the media removed and the cells resuspended in 200μl of 10mM EDTA. 20μl were removed to a new flat bottom 96 well plate containing 180μl of 10mM EDTA to create the "nonfloc OD" plate. The orginal plate was then spun down again, and the supernatant removed. The cells were resuspended in 180μl of flocculation buffer (0.51g/L CaSO4, 6.8g/L NaOAc, 4.05g/L acetic acid, 666mM Mannose) and shaken for 2 minutes at 150 rpm. Using a depth guide, 50μl of cells from 3mm below the surface were removed to a new 96 well plate containing 150μl of 0.1M EDTA to create the "floc OD" plate. OD_{600} for the nonfloc OD and floc OD plates were read. The following equation was used to calculate %flocculation from the OD_{600} readings:

$$\%floc = (nonfloc\ OD - floc\ OD)/nonfloc\ OD \times 100$$

Three biological replicates for each plate were averaged and strains with flocculation levels greater than two standard deviations away from the mean were called as hyperflocculant.

qPCR

Total RNA was obtained by standard acid phenol extraction from 2 ml of culture. The Qiagen QuantiTect Reverse Transcription Kit was used to remove residual genomic DNA and reverse transcribe the RNA templates to generate cDNAs. Aliquots of cDNA were used in Real Time PCR analyses with reagent from Applied Biosystems and the ABI7500 real-time PCR system.

Immunofluorescence

Cells from liquid cultures were washed washed twice with PBS, and resuspended i n 50μl PBS containing 0.25μl Alexafluor 488-conjugated anti-hemaglutinin antibody (Molecular Probes A-21287). Cells were incubated 30 minutes at 4C and washed three times in PBS and imaged with the Nikon Eclipse TE2000-S.

Figure A1-1 | **Flocculation and agar adhesion properties of S288c^{FLO8} and Sigma.** (A) Flocculation of Sigma, S288c, and S288c^{FLO8} strains correlates well with the known expression of *FLO1* solely in S288c^{FLO8}. Shown are overnight cultures grown to saturation in YPD and allowed to sediment for 5 minutes. All cultures are at approximately the same density but S288c^{FLO8} has sedimented to the bottom of the tube. (B) Agar adhesion of Sigma, S288c or S288c^{FLO8} strains correlates well with the known expression of *FLO11* in both Sigma and S288c^{FLO8} but not in S288c. Plate washing was performed after 3 days of growth at 30C.

1 - Sigma
2 - S288c
3 - S288c^{FLO8}
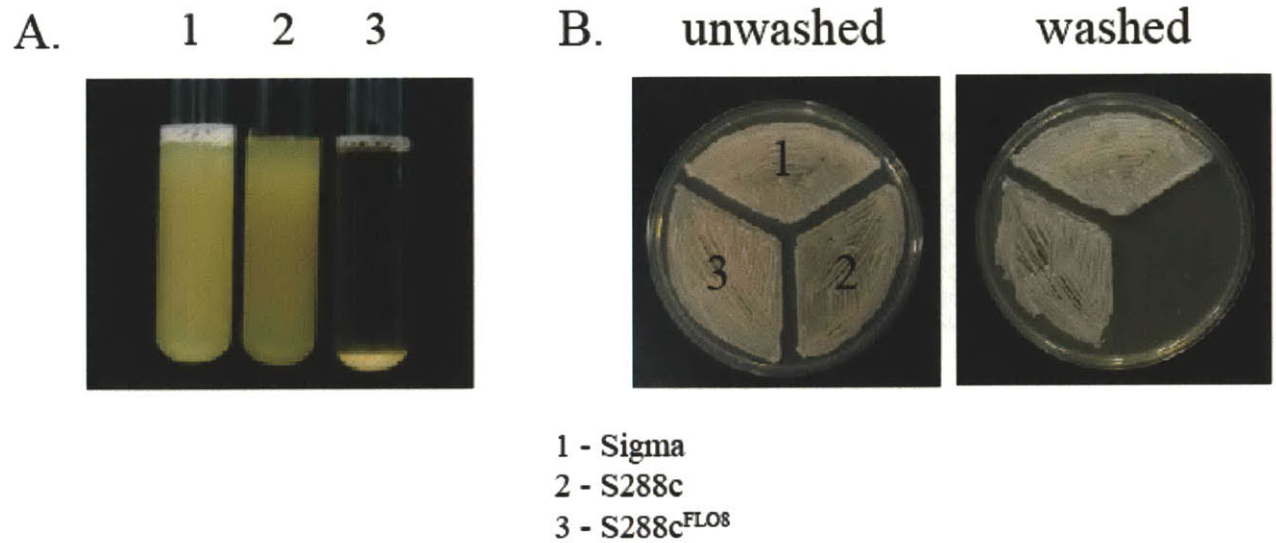
Figure A1-2 | **Genome-wide adhesion and flocculation screens examining S288c^{FLO8} agar adhesion and flocculation.** (A) Plate 47 from the agar adhesion screen is shown, with the S288c progenitor and S288c^{FLO8} as controls. (B) Results from screening plate 5 for flocculation. Each of the 95 strains is plotted as a ratio of their flocculant reading versus their non-flocculant reading and ranked from lowest to highest level of flocculation. The reading for *gas1Δ* is highlighted in green. (C) Venn diagram comparing the results from the S288c^{FLO8} adhesion and flocculation screens.

A.

unwashed

S288c *flo8–*
S288c *FLO8+*

washed

B.

Plate 5

C.

Figure A1-3 | **gas1Δ has contrasting effects on adhesion and flocculation.** (A) Plate washing assays reveal that deletion of *GAS1* in either S288c^{FLO8} or Sigma causes an adhesion defect with the effect being more pronounced in S288c^{FLO8}. (B) Consistent with the flocculation screen, S288c^{FLO8} *gas1Δ* is more flocculant than wild-type. Flocculation plotted as the mean of a ratio of the flocculant reading versus the non-flocculant reading ±SD.

A.

B.
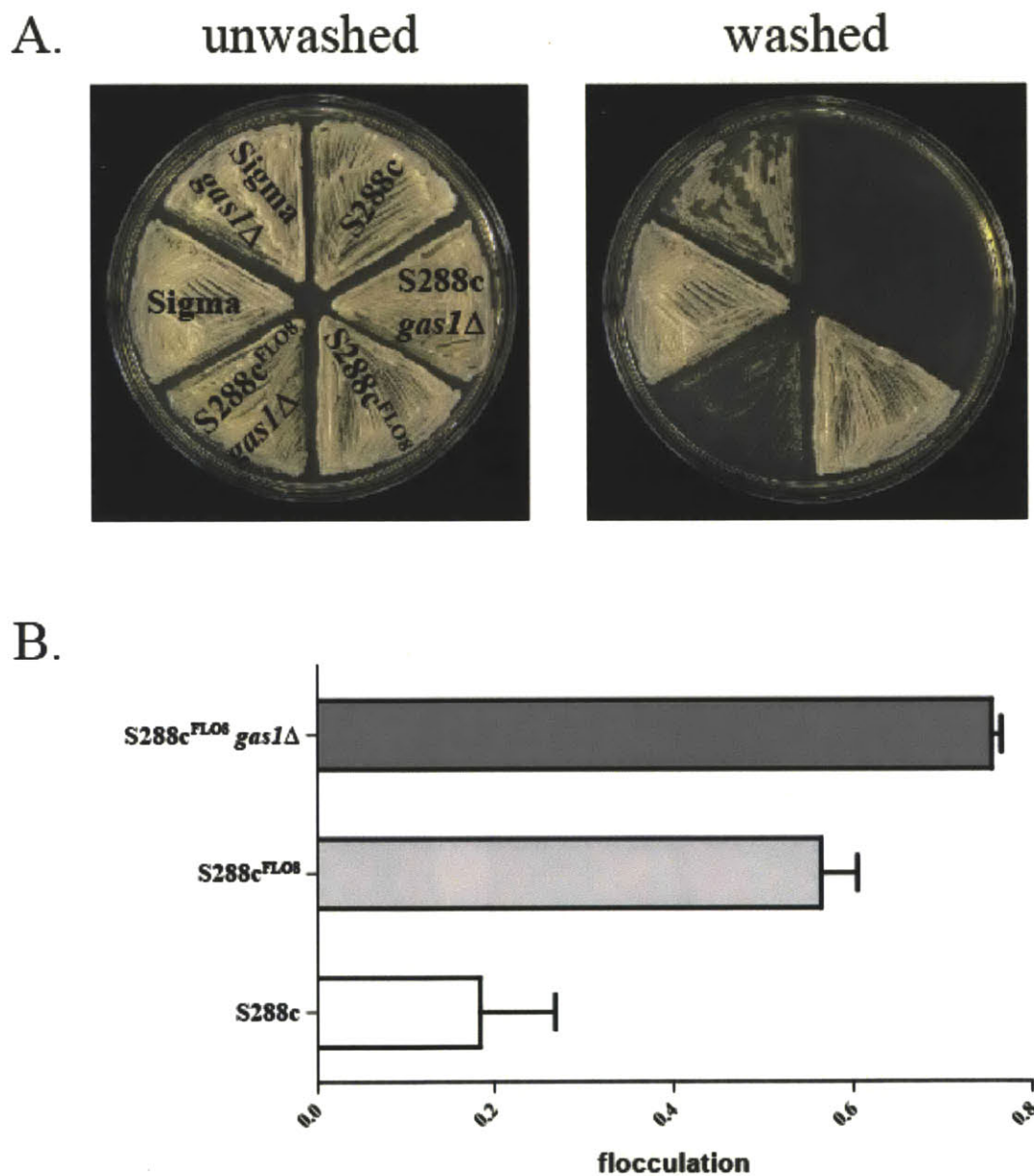
Figure A1-4 | *gas1Δ* alters yeast adhesive properties through multiple mechanisms. (A) qPCR of *FLO11* transcript levels performed on S288c^FLO8 and Sigma strains that were wild-type or carrying a *gas1Δ* shows that *gas1Δ* causes a decrease in *FLO11* levels independent of strain background. Mean *FLO11* levels from 3 biological replicates normalized to *ACT1* ±SD. (B) qPCR of *FLO1* transcript levels performed on S288c^FLO8 and Sigma strains that were wild-type or carrying a *gas1Δ* shows that *FLO1* levels are unaffected by deletion of *GAS1*. Mean *FLO1* levels from 3 biological replicates normalized to *ACT1* ±SD. (C) A plate washing assay of an S288c strain with the *FLO11* gene under the control of the constitutive *TEF* promoter shows that deleting *GAS1* can cause a loss of adhesion that is independent of *FLO11* transcription initiation.
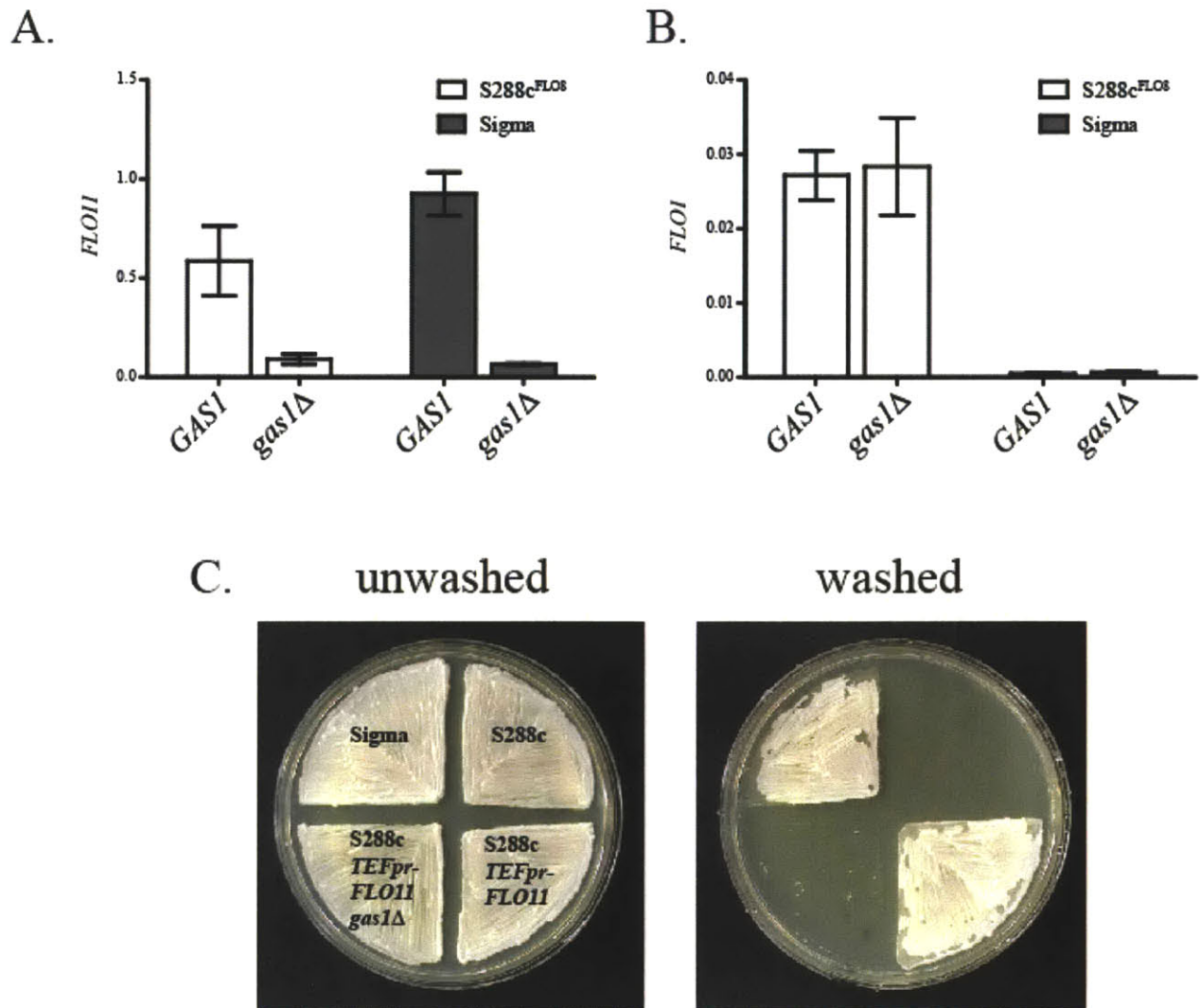
Figure A1-5 | *gas1Δ* **increases the epigenetic silencing of *FLO11* promoter.** Increased silencing at the *FLO11* promoter can be seen as a change in the proportion of cells that express different transcriptional fusions with the *FLO11* promoter. (A) 5-Fold serial dilutions of *FLO11pr-URA3* strains plated onto selection media (-Ura and 5-FOA) demonstrate that *gas1Δ* increases the proportion of 5-FOA resistant cells. (B) Fluorescent microscopy of *FLO11pr-GFP* strains shows that deletion of *GAS1* in both Sigma and S288c[FLO8] results in a decrease in the proportion of cells expressing GFP without a decrease in the maximum potential fluorescence.
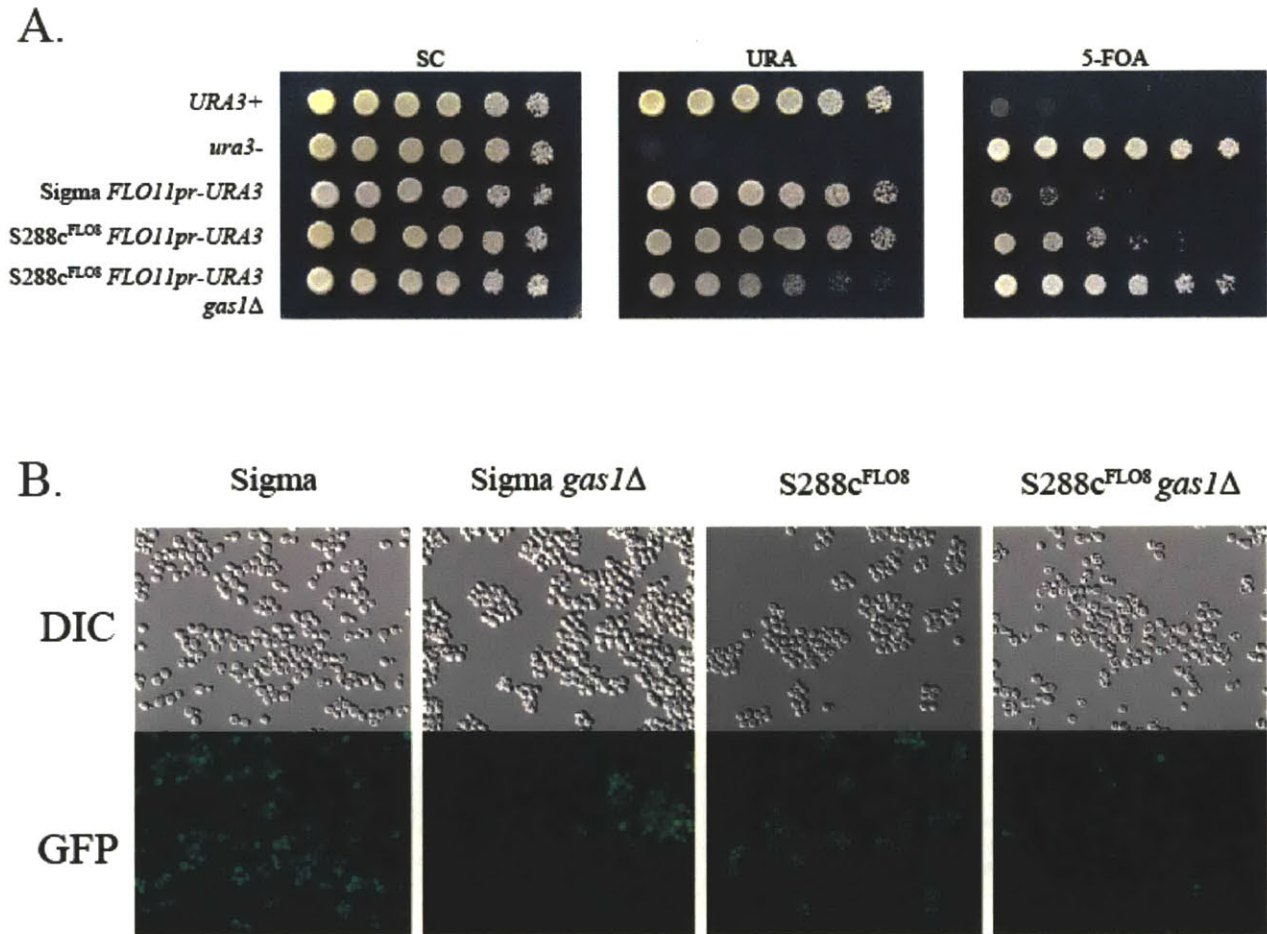
Figure A1-6 | **Zymolyase treatment phenocopies the *gas1Δ*.** S288c[FLO8] *FLO11pr-GFP* cells were incubated with or without zymolyase for half an hour at 30°C. Fluorescent microscopy shows a decrease in the number of GFP expressing cells upon zymolyase treatment, similar to what is seen in a *gas1Δ* strain.
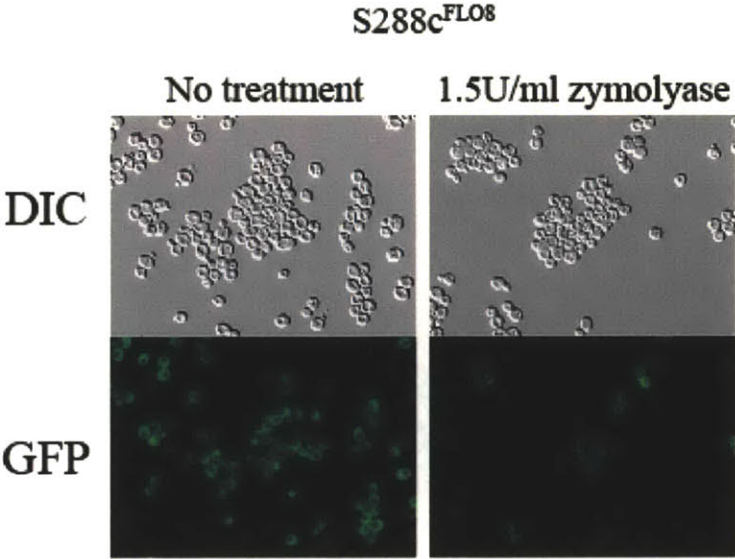
S288c[FLO8]

Figure A1-7 | **The SWR1 complex has a strain specific effect on adhesion.** (A) S288c[FLO8] strains deleted for *SWC3*, *SWR1*, or *HTZ1* have an adhesion defect as seen by a plate washing assay. These same deletions in Sigma have no adhesion defect. Agar adhesion assays performed on S288c strains (right half of the plate), and Sigma strains (left half of the plate). The same plate is shown before and after washing.
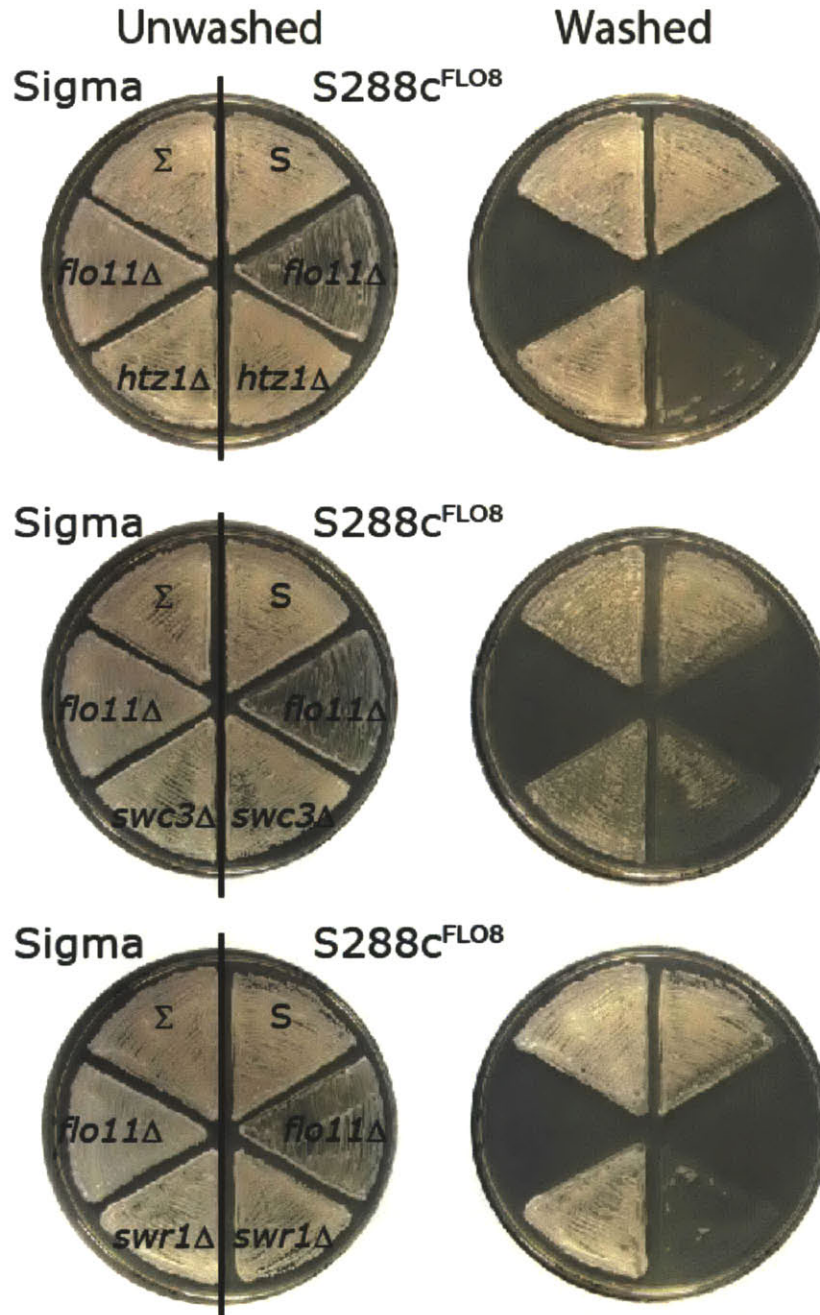
Figure A1-8 | **Deletion of components of the Swr1 does not affect *FLO11* transcript levels.**
FLO11 transcript levels examined in either *swr1Δ* (A) or *htz1Δ* (B) strains of S288c^{FLO8}
demonstrate that the adhesion defect is not due to a decrease in the level of *FLO11* transcript.
(C) Fluorescent microscopy of S288c^{FLO8} *FLO11pr-GFP* strains show that the deletion of *HTZ1*
does not lower transcription from the *FLO11* promoter.
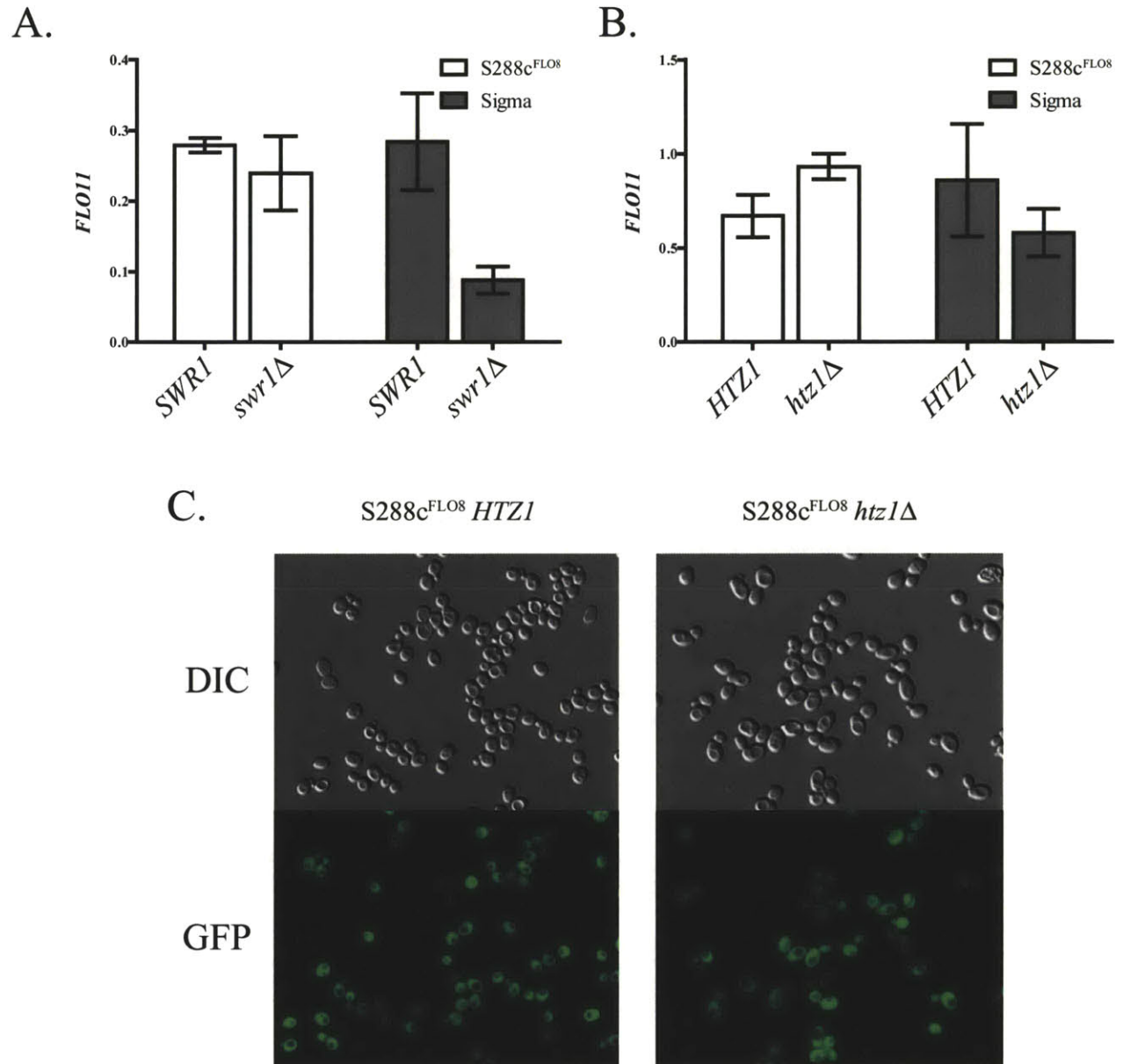
Figure A1-9 | *htz1Δ* does not affect Flo11p expression and localization.

Immunofluorescence microscopy examining *FLO11*-HA localization shows that deletion of *HTZ1* does not alter the expression or localization of *FLO11*-HA in either Sigma or S288c^{FLO8}.
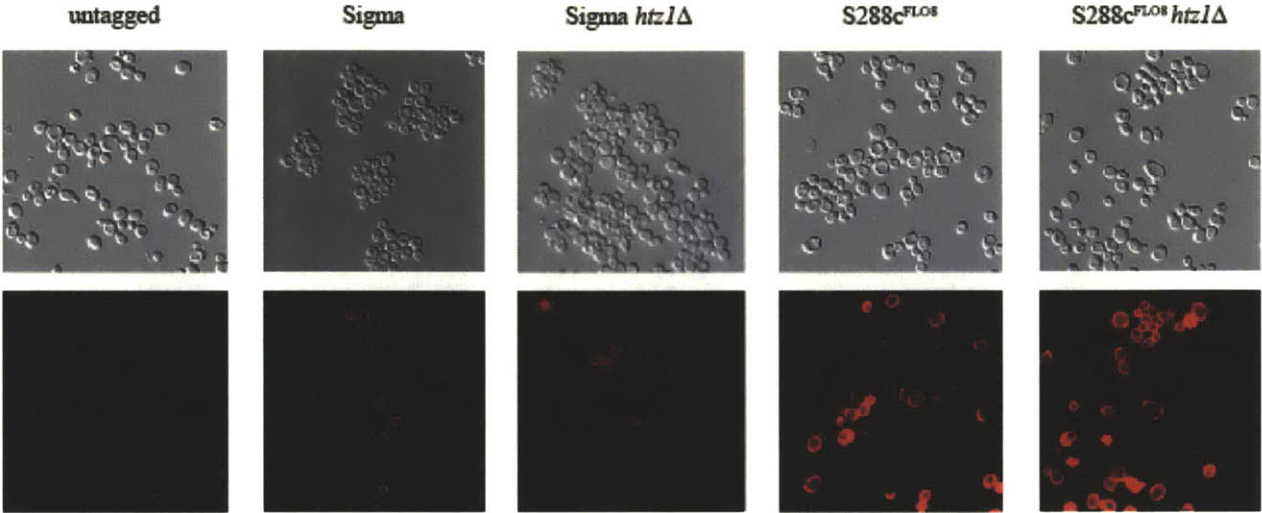
Figure A1-10 | **Altered MAPK signaling:** *HOG1* **and** *FUS3* **strain specific regulation of** *FLO11* **expression.** *hog1*Δ and *fus3*Δ affect (A) agar adhesion and (B) *FLO11* transcript levels in S288c^FLO8 but not in the Sigma strain, and the double mutant has a larger affect in S288c^FLO8 than either single mutant. qPCR data shown as the mean *FLO11* levels from 3 biological replicates normalized to *ACT1* ±SD.
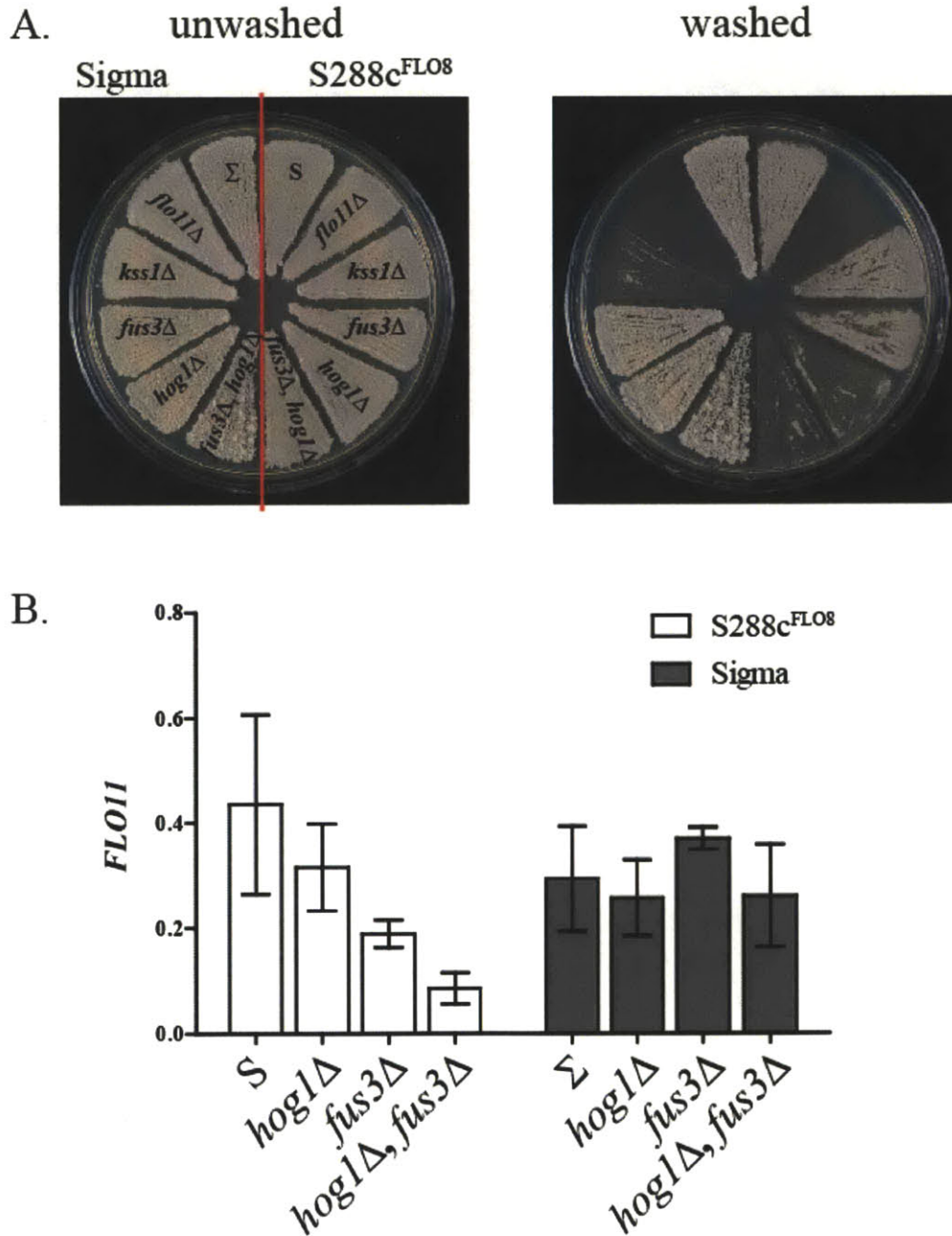
Table A1-1 | **GO term enrichment of S288c^{FLO8} adhesion modifiers.** Partial list of GO terms enriched in each phenotypic category for the S288c^{FLO8} flocculation and agar adhesion screens.

GO Term	P-value	Genes
Nonflocculant		
GO:0009408 response to heat	0.000666	WSC3 HSP104 GAC1 SDS3 WSC2 HOG1 SIN3 DEP1 RPB4 PBS2
GO:0031939 negative regulation of chromatin silencing at telomere	0.00244	SIF2 SDS3 RXT2 CTI6 SIN3 DEP1
Ahs-		
GO:0016568 chromatin modification	1.3E-09	ASF1 SNF5 SNF2 SWI3 SNF11 DEP1 HTZ1 RAD6 VPS72 SWC5 SWR1 RTF1 ASH1 YPL216W SDC1 UME6 SET3 UTH1 LDB7 RSC2 CBF1 RTT102 RXT2 BRE1 NGG1 HSL7 CTI6 VPS71 SIN3 SGF73 SPT3 YAF9 NPL6 RSC1 SPT8 ACS1
GO:0043486 histone exchange	0.00708	ASF1 VPS72 SWC5 SWR1 VPS71 YAF9
GO:0030447 filamentous growth	0.0084	RIM8 STE20 PTP1 SEC66 ASH1 TPK2 UME6 GPA2 MSS11 PHD1 RIM20 GPR1 ASC1 RXT2 RIM21 SPT3 DFG16
Hfl+		
no biological process		
Increased adhesion		
GO:0007120 axial cellular bud site selection	0.000529	BUD3 RSR1 ERV14 BUD2 AXL1

Table A1-2 | **Genes identified in the Hfl^{+} screen that are involved with cell wall properites.** A partial list of the deletions resulting that result in a hyper flocculant phenotype. Adhesion indicates if the deletion was also obtained from the Ahs^{-} screen. (n.c.) no change

Systematic ID	Gene	Description	Adhesion
YBL082C	*ALG3*	Alpha-1,3-mannosyltransferase	Decreased
YBR023C	*CHS3*	Synthesis of the majority of the cell well chitin	Decreased
YMR307W	*GAS1*	Beta-1,3-glucanosyltransferase	Decreased
YJL062W	*LAS21*	Synthesis of GPI core structure	Decreased
YJL183W	*MNN11*	Subunit of golgi mannosyltransferase complex	Decreased
YJL176C	*WSC1*	Sensor-transducer of the PKC-MPK1 kinase pathway	Decreased
YJL099W	*CHS6*	Mediates export of specific cargo proteins from the Golgi to plasma membrane	n.c.
YKL096W	*CWP1*	Cell wall mannoprotein	n.c.
YER124C	*DSE1*	Daughter cell-specific protein	n.c.
YJR075W	*HOC1*	Alpha-1,6-mannosyltransferase involved in cell wall mannan biosynthesis	n.c.
YDL240W	*LRG1*	Involved in the Pkc1p-mediated signaling pathway	n.c.
YGL028C	*SCW11*	Cell wall protein with similarity to glucanases	n.c.
YBL061C	*SKT5*	Activator of Chs3p (chitin synthase III)	n.c.
YDR293C	*SSD1*	Protein with a role in maintenance of cellular integrity	n.c.
YBR067C	*TIP1*	Major cell wall mannoprotein with possible lipase activity	n.c.
YLR425W	*TUS1*	GEF that functions to modulate Rho1p activity	n.c.

Table A1-3 | **List of strains used in this study**

Strain	Genotype	Source
BY4741	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, flo8-1*	Brachmann et al. (1998)
BY4742	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0,* lys2Δ0, *flo8-1*	Brachmann et al. (1998)
yBC9	S288c, MATx, *his3Δ1, leu2Δ0, ura3Δ0,* lys2Δ0, *FLO8*	this study
yBC37	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8*	this study
yBC28	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, FLO11pr-GFP::URA3*	this study
yBC46	S288c, MATx, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, FLO11pr-URA3*	this study
yBC160	S288c, MATa, *his3Δ1, ura3Δ0, FLO8, 3xHA-FLO11*	this study
yBC40	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, gas1Δ::KanMX4*	this study
yBC44	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, gas1Δ::KanMX4, FLO11pr-GFP::URA3*	this study
yBC58	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, gas1Δ::KanMX4, FLO11pr-URA3*	this study
yBC203	S288c, MATx, *his3Δ1, leu2Δ0, ura3Δ0,* lys2Δ0, *flo8-1, TEFpr-FLO11*	this study
yBC204	S288c, MATx, *his3Δ1, leu2Δ0, ura3Δ0,* lys2Δ0, *flo8-1, TEFpr-FLO11, gas1Δ::KanMX4*	this study
yBC01E5	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, htz1Δ::KanMX4*	this study
yBC02B6	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, swr1Δ::KanMX4*	this study
yBC02E3	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, swc3Δ::KanMX4*	this study
yBC147	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, htz1Δ::KanMX4, FLO11pr-GFP::URA3*	this study
yBC167	S288c, MATa, *his3Δ1, ura3Δ0, FLO8, 3xHA-FLO11, htz1Δ::KanMX4*	this study
yBC06A10	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, tec1Δ*::KanMX4	this study
yBC06E3	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, fus3Δ*::KanMX4	this study
yBC06D1	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, hog1Δ*::KanMX4	this study
yBC07A1	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO8, fus3Δ*::KanMX4, *hog1Δ*::HygMX4	this study
yBC0197	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO11pr-GFP*	this study
yBC08B11	S288c, MATa, *his3Δ1, leu2Δ0, ura3Δ0, met15Δ0, FLO11pr-GFP, tec1Δ::KanMX4*	this study
10560-6B	Sigma, MATx, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*	Fink

		Collection
yBC172	Sigma, MATa, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*	this study
L8225	Sigma, MATx, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *FLO11pr-GFP::URA3*	Fink Collection
L8383	Sigma, MATx, *leu2*::hisG, *ura3-52*, *FLO11pr-URA3*	Fink Collection
L7025	Sigma, MATx, *leu2*::hisG, *ura3-52*, *3xHA-FLO11*	Fink Collection
L8061	Sigma, MATx, *his3*::hisG, *leu2*::hisG, *ura3-52*, *gas1Δ::KanMX4*	Fink Collection
L8238	Sigma, MATx, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *FLO11pr-GFP::URA3, gas1Δ::KanMX4*	Fink Collection
yBC01B1	Sigma, MATx, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *htz1Δ::KanMX4*	this study
yBC02E9	Sigma, MATx, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *swr1Δ::KanMX4*	this study
yBC02C9	Sigma, MATx, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *swc3Δ::KanMX4*	this study
yBC204	Sigma, MATx, *leu2*::hisG, *ura3-52*, *3xHA-FLO11*, *htz1Δ::KanMX4*	this study
yBC06C9	Sigma, MATa, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *tec1Δ*::KanMX4	this study
yBC06E2	Sigma, MATa, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *fus3Δ*::KanMX4	this study
yBC07E7	Sigma, MATa, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *hog1Δ*::KanMX4	this study
yBC07G1	Sigma, MATa, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *fus3Δ::KanMX4, hog1Δ::HygMX4*	this study
yBC198	Sigma, MATx, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *FLO11pr-GFP*	this study
yBC08E11	Sigma, MATx, *his3*::hisG, *leu2*::hisG, trp1::hisG, *ura3-52*, *FLO11pr-GFP, tec1Δ::KanMX4*	this study
yBC09F11	Sigma/Sigma, MATa/x, *his3*::hisG/his3::hisG, *leu2*::hisG/leu2::hisG, trp1::hisG/trp1::hisG, *ura3-52*/ura3-52, *FLO11pr-GFP/FLO11pr-GFP, TEC/tec1Δ::KanMX4*	this study
yBC09H1	Sigma/S288c, MATa/x, *his3*::hisG/his3::hisG, *leu2*::hisG/leu2::hisG, trp1::hisG/trp1::hisG, *ura3-52*/ura3-52, *FLO11pr-GFP/FLO11pr-GFP, tec1Δ::KanMX4/tec1Δ::KanMX4*	this study

Table 3 | **List of oligonucleotides used in this study**

Name	Sequence (5' to 3')	Description
BCP46	GGAAACAAGCTGAGCTGGAC	Flanking *TEC1*
BCP47	TCGTGGTTTCATCCAAGTGA	Flanking *TEC1*
BCP187	TTAAAACATCACGCGATCCA	Flanking *FUS3*
BCP188	TTTTATACGTCCGCGTCCTC	Flanking *FUS3*
BCP191	CCCAAGCGAGACCTAGAGTG	Flanking *STE12*
BCP192	GAACATCGATGCCTTCACCT	Flanking *STE12*
BCP195	AAGTGATTCGTGGGGTAACG	Flanking *STE7*
BCP196	TGGGTTATTAATCGCCTTCG	Flanking *STE7*
BCP199	ATTCTCGCCCAACTTTTCCT	Flanking *STE11*
BCP200	TCTTCGTGCTTCCATCTGTG	Flanking *STE11*
BCP185	GCGCAAGTTGTTAGGAAAGC	Flanking *HOG1*
BCP186	CGCCATAAGTGACGGTTCTT	Flanking *HOG1*
FLO11 FW	cacttttgaagtttatgccacacaag	*FLO11* qPCR
FLO11 RV	cttgcatattgagcggcactac	*FLO11* qPCR
ACTI FW	ctccaccactgctgaaagagaa	*ACT1* qPCR
ACTI RV	ccaaggcgacgtaacatagtttt	*ACT1* qPCR

References

AHN, S. H., A. ACURIO and S. J. KRON, 1999 Regulation of G2/M progression by the STE mitogen-activated protein kinase pathway in budding yeast filamentous growth. Mol Biol Cell **10:** 3301-3316.

BESTER, M. C., I. S. PRETORIUS and F. F. BAUER, 2006 The regulation of Saccharomyces cerevisiae FLO gene expression and Ca2+ -dependent flocculation by Flo8p and Mss11p. Curr Genet **49:** 375-383.

BRACHMANN, C. B., A. DAVIES, G. J. COST, E. CAPUTO, J. LI et al., 1998 Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast **14:** 115-132.

BUMGARNER, S. L., R. D. DOWELL, P. GRISAFI, D. K. GIFFORD and G. R. FINK, 2009 Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. Proc Natl Acad Sci U S A **106:** 18321-18326.

CHANDARLAPATY, S., and B. ERREDE, 1998 Ash1, a daughter cell-specific protein, is required for pseudohyphal growth of Saccharomyces cerevisiae. Mol Cell Biol **18:** 2884-2891.

CHEN, R. E., and J. THORNER, 2007 Function and regulation in MAPK signaling pathways: lessons learned from the yeast Saccharomyces cerevisiae. Biochim Biophys Acta **1773:** 1311-1340.

CHEN, R. E., and J. THORNER, 2010 Systematic epistasis analysis of the contributions of protein kinase A- and mitogen-activated protein kinase-dependent signaling to nutrient limitation-evoked responses in the yeast Saccharomyces cerevisiae. Genetics **185:** 855-870.

DAGUE, E., R. BITAR, H. RANCHON, F. DURAND, H. M. YKEN et al., 2010 An atomic force microscopy analysis of yeast mutants defective in cell wall architecture. Yeast **27:** 673-684.

DIETVORST, J., and A. BRANDT, 2008 Flocculation in Saccharomyces cerevisiae is repressed by the COMPASS methylation complex during high-gravity fermentation. Yeast **25:** 891-901.

FURUKAWA, K., F. SIDOUX-WALTER and S. HOHMANN, 2009 Expression of the yeast aquaporin Aqy2 affects cell surface properties under the control of osmoregulatory and morphogenic signalling pathways. Mol Microbiol **74:** 1272-1286.

GIMENO, C. J., P. O. LJUNGDAHL, C. A. STYLES and G. R. FINK, 1992 Unipolar cell divisions in the yeast S. cerevisiae lead to filamentous growth: regulation by starvation and RAS. Cell **68:** 1077-1090.

GUO, B., C. A. STYLES, Q. FENG and G. R. FINK, 2000 A Saccharomyces gene family involved in invasive growth, cell-cell adhesion, and mating. Proc Natl Acad Sci U S A **97:** 12158-12163.

GUTHRIE, C., and G. FINK, 2002 Guide to yeast genetics and molecular and cellular biology.

HALME, A., S. BUMGARNER, C. STYLES and G. R. FINK, 2004 Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. Cell **116:** 405-415.

KOBAYASHI, O., H. YOSHIMOTO and H. SONE, 1999 Analysis of the genes activated by the FLO8 gene in Saccharomyces cerevisiae. Curr Genet **36:** 256-261.

KOBOR, M. S., S. VENKATASUBRAHMANYAM, M. D. MENEGHINI, J. W. GIN, J. L. JENNINGS et al., 2004 A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. PLoS Biol **2:** E131.

KOCH, M. R., and L. PILLUS, 2009 The glucanosyltransferase Gas1 functions in transcriptional silencing. Proc Natl Acad Sci U S A **106:** 11224-11229.

KROGAN, N. J., M. C. KEOGH, N. DATTA, C. SAWA, O. W. RYAN et al., 2003 A Snf2 family ATPase complex required for recruitment of the histone H2A variant Htz1. Mol Cell **12:** 1565-1576.

KRON, S. J., C. A. STYLES and G. R. FINK, 1994 Symmetric cell division in pseudohyphae of the yeast Saccharomyces cerevisiae. Mol Biol Cell **5:** 1003-1022.

KUCHIN, S., V. K. VYAS and M. CARLSON, 2002 Snf1 protein kinase and the repressors Nrg1 and Nrg2 regulate FLO11, haploid invasive growth, and diploid pseudohyphal differentiation. Mol Cell Biol **22:** 3994-4000.

LIU, H., C. A. STYLES and G. R. FINK, 1993 Elements of the yeast pheromone response pathway required for filamentous growth of diploids. Science **262:** 1741-1744.

LIU, H., C. A. STYLES and G. R. FINK, 1996 Saccharomyces cerevisiae S288C has a mutation in FLO8, a gene required for filamentous growth. Genetics **144:** 967-978.

LO, H. J., J. R. KOHLER, B. DIDOMENICO, D. LOEBENBERG, A. CACCIAPUOTI *et al.*, 1997 Nonfilamentous C. albicans mutants are avirulent. Cell **90:** 939-949.

LO, W. S., and A. M. DRANGINIS, 1998 The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by Saccharomyces cerevisiae. Mol Biol Cell **9:** 161-171.

MADHANI, H. D., and G. R. FINK, 1997 Combinatorial control required for the specificity of yeast MAPK signaling. Science **275:** 1314-1317.

MADHANI, H. D., T. GALITSKI, E. S. LANDER and G. R. FINK, 1999 Effectors of a developmental mitogen-activated protein kinase cascade revealed by expression signatures of signaling mutants. Proc Natl Acad Sci U S A **96:** 12530-12535.

MADHANI, H. D., C. A. STYLES and G. R. FINK, 1997 MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation. Cell **91:** 673-684.

MALCHER, M., S. SCHLADEBECK and H. U. MOSCH, 2011 The Yak1 protein kinase lies at the center of a regulatory cascade affecting adhesive growth and stress resistance in Saccharomyces cerevisiae. Genetics **187:** 717-730.

MENEGHINI, M. D., M. WU and H. D. MADHANI, 2003 Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell **112:** 725-736.

MIZUGUCHI, G., X. SHEN, J. LANDRY, W. H. WU, S. SEN *et al.*, 2004 ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. Science **303:** 343-348.

MOSCH, H. U., E. KUBLER, S. KRAPPMANN, G. R. FINK and G. H. BRAUS, 1999 Crosstalk between the Ras2p-controlled mitogen-activated protein kinase and cAMP pathways during invasive growth of Saccharomyces cerevisiae. Mol Biol Cell **10:** 1325-1335.

MOUYNA, I., T. FONTAINE, M. VAI, M. MONOD, W. A. FONZI *et al.*, 2000 Glycosylphosphatidylinositol-anchored glucanosyltransferases play an active role in the biosynthesis of the fungal cell wall. J Biol Chem **275:** 14882-14889.

O'ROURKE, S. M., and I. HERSKOWITZ, 1998 The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in Saccharomyces cerevisiae. Genes Dev **12:** 2874-2886.

OCTAVIO, L. M., K. GEDEON and N. MAHESHRI, 2009 Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. PLoS Genet **5:** e1000673.

PALECEK, S. P., A. S. PARIKH and S. J. KRON, 2000 Genetic analysis reveals that FLO11 upregulation and cell polarization independently regulate invasive growth in Saccharomyces cerevisiae. Genetics **156:** 1005-1023.

PAN, X., T. HARASHIMA and J. HEITMAN, 2000 Signal transduction cascades regulating pseudohyphal differentiation of Saccharomyces cerevisiae. Curr Opin Microbiol **3:** 567-572.

PAN, X., and J. HEITMAN, 1999 Cyclic AMP-dependent protein kinase regulates pseudohyphal differentiation in Saccharomyces cerevisiae. Mol Cell Biol **19:** 4874-4887.

POPOLO, L., M. VAI, E. GATTI, S. PORELLO, P. BONFANTE *et al.*, 1993 Physiological analysis of mutants indicates involvement of the Saccharomyces cerevisiae GPI-anchored protein gp115 in morphogenesis and cell separation. J Bacteriol **175:** 1879-1885.

RAM, A. F., J. C. KAPTEYN, R. C. MONTIJN, L. H. CARO, J. E. DOUWES *et al.*, 1998 Loss of the plasma membrane-bound protein Gas1p in Saccharomyces cerevisiae results in the release of beta1,3-glucan into the medium and induces a compensation mechanism to ensure cell wall integrity. J Bacteriol **180:** 1418-1424.

RAM, A. F., A. WOLTERS, R. TEN HOOPEN and F. M. KLIS, 1994 A new approach for isolating cell wall mutants in Saccharomyces cerevisiae by screening for hypersensitivity to calcofluor white. Yeast **10:** 1019-1030.

ROBERTS, R. L., and G. R. FINK, 1994 Elements of a single MAP kinase cascade in Saccharomyces cerevisiae mediate two developmental programs in the same cell type: mating and invasive growth. Genes Dev **8:** 2974-2985.

ROBERTS, R. L., H. U. MOSCH and G. R. FINK, 1997 14-3-3 proteins are essential for RAS/MAPK cascade signaling during pseudohyphal development in S. cerevisiae. Cell **89:** 1055-1065.

RUPP, S., E. SUMMERS, H. J. LO, H. MADHANI and G. FINK, 1999 MAP kinase and cAMP filamentation signaling pathways converge on the unusually large promoter of the yeast FLO11 gene. EMBO J **18:** 1257-1269.

SANTISTEBAN, M. S., T. KALASHNIKOVA and M. M. SMITH, 2000 Histone H2A.Z regulats transcription and is partially redundant with nucleosome remodeling complexes. Cell **103:** 411-422.

SCHWARTZ, M. A., and H. D. MADHANI, 2004 Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae. Annu Rev Genet **38:** 725-748.

SCHWARTZ, M. A., and H. D. MADHANI, 2006 Control of MAPK signaling specificity by a conserved residue in the MEK-binding domain of the yeast scaffold protein Ste5. Curr Genet **49:** 351-363.

STRATFORD, M., 1989 Evidence for two mechanisms of flocculation in Saccharomyces cerevisiae. Yeast **5 Spec No:** S441-445.

STRATFORD, M., and S. ASSINDER, 1991 Yeast flocculation: Flo1 and NewFlo phenotypes and receptor structure. Yeast **7:** 559-574.

VAN DYK, D., I. S. PRETORIUS and F. F. BAUER, 2005 Mss11p is a central element of the regulatory network that controls FLO11 expression and invasive growth in Saccharomyces cerevisiae. Genetics **169:** 91-106.

VENKATASUBRAHMANYAM, S., W. W. HWANG, M. D. MENEGHINI, A. H. TONG and H. D. MADHANI, 2007 Genome-wide, as opposed to local, antisilencing is mediated redundantly by the euchromatic factors Set1 and H2A.Z. Proc Natl Acad Sci U S A **104:** 16609-16614.

VERSTREPEN, K. J., G. DERDELINCKX, H. VERACHTERT and F. R. DELVAUX, 2003 Yeast flocculation: what brewers should know. Appl Microbiol Biotechnol **61:** 197-205.

VERSTREPEN, K. J., and F. M. KLIS, 2006 Flocculation, adhesion and biofilm formation in yeasts. Mol Microbiol **60:** 5-15.

VINOD, P. K., and K. V. VENKATESH, 2007 Specificity of MAPK signaling towards FLO11 expression is established by crosstalk from cAMP pathway. Syst Synth Biol **1:** 99-108.

VOYNOV, V., K. J. VERSTREPEN, A. JANSEN, V. M. RUNNER, S. BURATOWSKI *et al.*, 2006 Genes with internal repeats require the THO complex for transcription. Proc Natl Acad Sci U S A **103:** 14423-14428.

VYAS, V. K., S. KUCHIN, C. D. BERKEY and M. CARLSON, 2003 Snf1 kinases with different beta-subunit isoforms play distinct roles in regulating haploid invasive growth. Mol Cell Biol **23:** 1341-1348.

WESTFALL, P. J., and J. THORNER, 2006 Analysis of mitogen-activated protein kinase signaling specificity in response to hyperosmotic stress: use of an analog-sensitive HOG1 allele. Eukaryot Cell **5:** 1215-1228.

WINZELER, E. A., B. LEE, J. H. MCCUSKER and R. W. DAVIS, 1999 Whole genome genetic-typing in yeast using high-density oligonucleotide arrays. Parasitology **118 Suppl:** S73-80.

WOLF, J. J., R. D. DOWELL, S. MAHONY, M. RABANI, D. K. GIFFORD *et al.*, 2010 Feed-forward regulation of a cell fate determinant by an RNA-binding protein generates asymmetry in yeast. Genetics **185:** 513-522.

Appendix A2

FASTER MT: Isolation of Pure Populations of *a* and α Ascospores from *Saccharomyces cerevisiae*

Brian L. Chin[1], Margaret A. Frizzell, William E. Timberlake, and Gerald R. Fink
Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge, MA 02142

[1]Present Address: Department of Systems Biology
Harvard Medical School
200 Longwood Ave. WAB 536
Boston, MA 02115

Running Title: Purification of Yeast Ascospores by Mating Type

Key Words: budding yeast, red fluorescent protein, *MATa*, fluorescence-activated cell sorting, hygromycin resistance, *BUD5-TAF2*

Corresponding Author: William E. Timberlake
Address: Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge, MA 02142
Phone: 617-642-8631
FAX: 617-258-9872
E-mail: wtimbe@wi.mit.edu

Abstract

The budding yeast *Saccharomyces cerevisiae* has many traits that make it useful for studies of quantitative inheritance. Genome-wide association studies (GWAS) and bulk segregant analyses (BSA) often serve as first steps toward identification of quantitative trait loci (QTL). These approaches benefit from having large numbers of ascospores pooled by mating type without contamination by vegetative cells. To this end, we inserted a gene encoding red fluorescent protein (RFP) into the *MATa* locus. RFP expression caused *MATa* and *a/α* diploid vegetative cells and *MATa* ascospores to fluoresce; *MATα* cells

without the gene did not fluoresce. Heterozygous diploids segregated fluorescent and non-fluorescent ascospores 2:2 in tetrads and bulk populations. The two populations of spores were separable by fluorescence-activated cell sorting (FACS) with little cross contamination or contamination with diploid vegetative cells. This approach, which we call *F*luorescent *As*cospore *T*echnique for *E*fficient *R*ecovery of *M*ating *T*ype (*FASTER MT*), should be applicable to laboratory, industrial, and undomesticated, strains.

Introduction

Mapping and identification of quantitative trait loci (QTL) is key to understanding complex traits in humans, animals, plants, and eukaryotic microorganisms (Lander and Schork, 1994; Darvasi, 1998; Gianni, 2009; Jing et al., 2010; Jimenez-Gomez et al., 2011). Such studies often utilize hundreds or even thousands of individuals (Churchill and Doerge, 1994) in order to detect associations or linkages between genetic markers, such as single nucleotide polymorphisms (SNPs), and traits of interest. *Saccharomyces cerevisiae* (yeast) is well suited for QTL mapping studies: Its facile genetic system, small genome size, and lack of extensive repeated DNA make it ideal for developing strategies to detect the many loci contributing to complex traits in eukaryotes. Further, precise control of the cellular environment when growing yeast minimizes non-genetic variability and thereby increases the ability to detect quantitative variation caused by genetic differences. The potential for yeast to help solve basic problems in quantitative genetics has been, for example, exploited in studies of sporulation (Deutschbauer and Davis, 2005), heat tolerance (Steinmetz et al., 2002), and chemical tolerance (Ehrenreich et al., 2010).

In yeast, meiotic segregants can be isolated by micromanipulation of individual tetrads to separate the four ascospores or as random spores, where ascus walls are enzymatically removed and the population of released spores is plated. Because tetrad analysis is time consuming and not automated it is ill suited to produce sufficient numbers of recombinant progeny for QTL studies. Isolation of large numbers of random spores without micromanipulation is straightforward, but has at least two technical shortcomings. First, a diploid culture subjected to meiosis-inducing conditions contains contaminating diploids that failed to undergo meiosis in addition to the desired haploid meiotic spores. Second, the population of haploid meiotic cells consists of equal numbers of the two mating types, which when plated could mate to form diploids. Without a method for removing diploids and separating haploids into *a* and *α* mating types the random spore population is not useful for QTL mapping. Thus, simple, rapid, and efficient methods for bulk isolation of pure ascospores sorted by mating type are needed.

Rapid separation of haploids and diploids has been accomplished by incorporation of genetic markers that allow for selection by 1) insertion of a gene-promoter construct expressed only in haploids of one mating type and 2) utilization of a recessive resistance marker (e.g. canavanine-resistance; Whelan et al., 1979) to select against diploids (Ehrenreich et al., 2010; Tong and Boone, 2007). Although effective, these approaches require introduction of engineered cassettes via multiple manipulations and entail selections that could bias some analyses. Further, they may not be applicable to wild strains, which are rich sources of quantitative variation but are diploid, often homothallic, and lack genetic markers needed for introduction of some engineered cassettes (Timberlake et al., 2011).

Thacker et al., 2011 demonstrated the feasibility of obtaining ascospore-autonomous expression of fluorescent protein constructs and used these to visualize meiotic events. Fluorescently tagged ascospores would be well suited for preparation of QTL mapping populations if expression of the tag could be limited to one mating type. The approach we describe here is based on integration of a red fluorescent protein (RFP) gene at the *MATa* locus with selection provided by a hygromycin-resistance gene so that the cassette can be introduced into any transformable, haploid or diploid, hygromycin-sensitive strain. *MATa* vegetative cells and ascospores thus tagged contain a visible marker useful for separation of cells by hand or fluorescence-activated cell sorting (FACS).

Materials and Methods

We used standard yeast molecular genetic techniques (Guthrie and Fink, 2004; Amberg, Burke, and Strathern, 2005) to obtain the *S. cerevisiae* Σ1278b (http://wiki.yeastgenome.org/index.php/History_of_Sigma) strains given in Table 1.

Plasmid pBC58 (Figure A2-1A) was constructed as follows: A *Bam*H1 fragment from plasmid yEpGAP-Cherry (Keppler-Ross, Noffz, and Dean, 2008) containing a yeast-optimized red fluorescent protein gene and promoter (*TDH3p$_{yEmRFP}$*) was cloned into pAG35 (Goldstein and McCusker, 1999). A PCR product (BCP538-539; Table 2) encompassing the RFP-hygMX genes and adding ~50 bp of homology at the 5' end of *MATa2* was used to direct integration at *MATa*. A second PCR product (BCP569-571; Table 2) spanning *MATa* and adding terminal *Stu*I sites was then made from genomic DNA and cloned into pCR TOPO2.1 (Invitrogen) to produce pBC58, which is available upon request.

Cells were examined with a 40x/0.75 M/N2 dry objective or 100x/1.30 H/N2 oil immersion objective at room temperature. Fluorescence was monitored at 590 nm with a G-2F/C blocking filter (Nikon).

FACS was performed with either a BD Biosciences FACS AriaIIU SORP or LSRII SORP with the 561 nm laser and 610/20 filter.

Growth curves were performed in microtiter plates with 150 µl of medium/well. Wells were inoculated with 10 µl of 1 OD$_{600}$/ml aqueous suspensions of cells. Plates were incubated at 30° and OD$_{600}$ measurements were taken at 30 min intervals after shaking for 15 sec.

Ascospores were isolated by scraping well-sporulated colonies from SM plates, suspending them in 1 ml PBS, and adding 1000 units of lyticase (Sigma-Aldrich). After incubation at 30° for 8 hr SDS was added to 1%. The ascospores were washed twice with 0.1% Tween-20, 5 mM EDTA, and suspended at ~10^9/ml.

Results and Discussion

Transformation of haploid *MATa* strains with the *Stu*I fragment of pBC58 (Figure A2-1A) resulted in the formation of hygromycin-resistant (hyg^R), pink colonies. The intensity of the color increased upon incubation at 4°. We crossed one transformant to produce heterozygous diploid ML1 (Table 1), whose color was approximately one-half as intense as that of the haploid. Figures A1-1B-D show that segregation of the marker in ML1 tetrads was 2 RFP$^+$:2 rfp$^-$. PCR analysis of transformants indicated that a single copy of RFP-hygMX had integrated at *MATa*. Moreover, mating type was completely linked to RFP in 20 tetrads. These results indicate that transformation was due to integration by homology at *MATa*.

Transformation of a diploid strain with the *Stu*I fragment also resulted in formation of hyg^R, pink colonies. Of these, ~10% were converted from *a/α* to *a/a* diploids as evidenced by acquisition of mating competence with a *MATα* tester lawn. This is predicted by transplacement of the *MATα* locus by the pBC58 *Stu*I fragment, which contains homologous sequences flanking *MAT* (*BUD5/TAF2*; Figure A2-1A). The ability to make RFP$^+$/rfp$^-$ diploids by transformation speeds up analysis, because strains can be sporulated without intermediate steps to obtain segregants. Further, Klar (1980) showed that *a/a* diploids could be induced to sporulate after transient mating with a *MATα* haploid containing a *kar1* mutation that interfered with karyogamy (Conde and Fink, 1976). This approach, which is expected to produce equal numbers of *MATa* spores containing and lacking the *RASTER* insert, could be used to obtain untagged *MATa* populations.

We subjected vegetative cells and ascospores to FACS to assess the feasibility of separating them by mating type. Figure A2-2A shows that control haploid cells (non-transformed or *MATα* derivatives of transformed diploids) and transformed haploids are separated by ~3 logs of intensity, whereas heterozygous diploids are intermediate. Gating permitted separation of the three classes: diploids and *MATa* and MATα haploids. Separation of ascospores is more relevant to most studies. Figure A2-2B shows that forward and side scatter analysis separated a crude ascospore preparation into four

populations, one of which contained equal numbers of individual fluorescent and non-fluorescent cells. Microscopic examination of these cells showed that they were un-aggregated ascospores. Figure A2-1C shows that this population could be sorted into non-overlapping, non-fluorescent and fluorescent sub-populations, present in equal proportions.

We tested each population for viability and cross-contamination ($MATa \rightarrow MAT\alpha$ and converse). Table 3 shows that spore viability was high (60-70%) even after the rigorous enzymatic and detergent treatments used to eliminate ascus walls and vegetative cells, and FACS. For the $MAT\alpha$ (non-fluorescent) population the contamination with hyg^R cells was <0.2%, which should be acceptable for most purposes. Moreover, as the contaminating cells, which we presume are the result of aggregation, are RFP$^+$, they can be removed without much effort after plating because the colonies are red. The $MATa$ (fluorescent) population was contaminated with ~0.2% of fluorescent diploid cells (Table 3). These could be removed by further enzymatic and detergent treatments.

These results lead to the following conclusions:

1. Large, pure populations of $MATa$ and $MAT\alpha$ spores can be obtained by FACS. These have high viability making them suitable for GWAS and BSA.

2. RFP is expressed at high enough levels to be detected visually in colonies. Therefore, as RFP and hyg^R are completely linked to $MATa$, haploid colonies can be separated into mating types by fluorescence or drug resistance.

3. The ability to use both $MATa$ and $MAT\alpha$ populations lacking the introduced marker provides a way to get around potential distortions arising from linkage of genes of interest to MAT.

Although deletion of $MATa2$ has been reported to have no effect on growth, mating, or sporulation (Dranginis, 1989), we assessed the growth characteristics and mating competence of some of our $MATa2$ transplacement strains. Figure A2-3 shows growth curves of strains ML1-4 (Table 1). ML1, a diploid containing the RFP cassette, and ML2, a related diploid lacking the cassette, had similar growth

profiles on either YPD or supplemented SD, although ML1 reproducibly grew a little slower. By contrast, ML3, a haploid containing the cassette, grew much more slowly and to a lower final OD in YPD than isogenic ML4 lacking the cassette. However, this difference was moderated and reversed in SD. These differences could be a consequence of the insertion of two strong promoters at *MATa2* and suggest that controlled measurements of growth rates (or other traits of interest) are required for strains containing the RFP cassette. Of course this caution applies to any strains carrying residual markers, selection cassettes, chromosome abnormalities, etc., introduced to facilitate QTL studies, because they might modify or bias traits of interest directly or indirectly.

We found that RFP strains mated as well as non-RFP strains in routine strain constructions. However, in a mating assay where congenic RFP$^+$ and rfp$^-$ strains were in competition for a common mating partner the RFP$^+$ strain mated somewhat less efficiently than the rfp$^-$ strain. This disadvantage decreased with increased mating time. Dranginis, 1989, reported that strains containing a complete deletion of *MATa2* had normal mating characteristics, but this conclusion was not based on the sensitive competitive assays employed here. Whatever the function of *MATa2* and the effect of the insertion, RFP strains in which it is disrupted mate well under the standard, non-competitive conditions used for strain construction.

These results lead to the following conclusions:

1. Integration of the RFP cassette at *MATa* does not influence growth rate on one medium, but does on another. Growth rates of selective markers should be assessed in QTL studies.

2. The RFP cassette does not interfere with standard genetic manipulations, but may reduce mating efficiency in more sensitive assays.

Summary: Integration of a cassette containing RFP and hyg^R into the *MATa* locus provides a simple, robust means for marking mating type so that *a* and *α* ascospores can be separated and purified by FACS. The fact that the cassette can be transformed into most haploid or diploid *S.*

cerevisiae strains without introduction of other mutations means that it should be useful for

studies of quantitative inheritance in laboratory, industrial, and wild strains. Moreover, it can

serve as a mating type indicator without compromising other genotypic or phenotypic features.

Figure A2-1 | **Transformation with the RFP Cassette**. (A) Plasmid pBC58. The RFP-hyg^R cassette was inserted between the first and second codons of *HMRa2*. The figure retains the *HMRa* notation, because the *MATa* sequence was first inferred from the sequence of the silenced locus. However, the cassette's homology extends to the flanking *TAF2* and *BUD5* genes so transformation with the *Stu*I fragment is directed to the *MAT* locus. (B), (C) Fluorescence phenotype of asci. The great majority of the intact asci we observed contained two fluorescent and two non-fluorescent spores. The RFP appeared to accumulate in vacuoles. (D) Growth of tetrads. Dissected tetrads were grown at 30° on YPD medium, incubated at 4° for several days to enhance fluorescence, and photographed under ambient light. Normal segregation of fluorescent ascospores shown in panels (B) and (C) was replicated in these and all other tetrads we observed. We confirmed that *a* mating type, fluorescence, and hygromycin-resistance were completely linked. By contrast, the variations in colony morphology shown in the figure were unlinked to fluorescence.
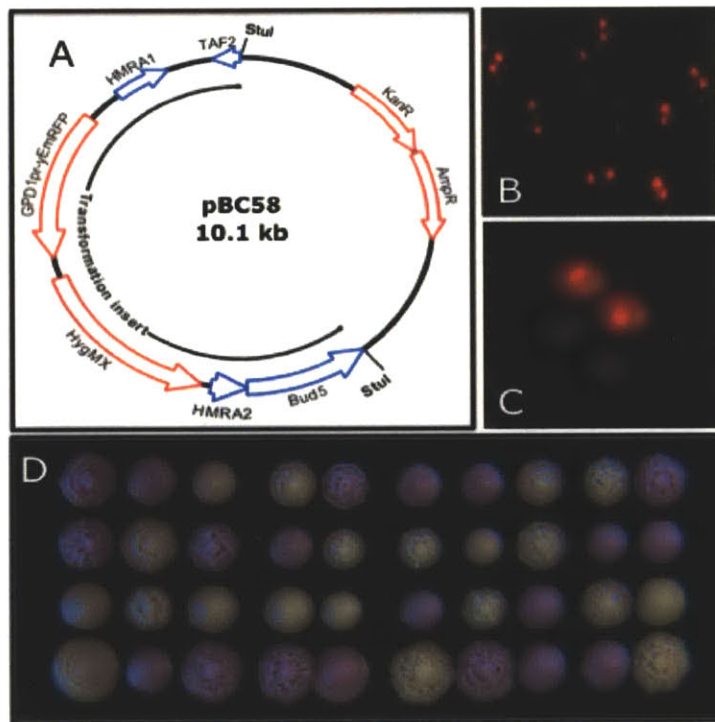
Figure A2-2 | **Fluorescence-activated cell sorting (FACS).** (A) Vegetative cells: upper panel--non-transformed *MATα* haploids; middle panel--transformed *MATa* haploids; lower panel--heterozygous diploid. (B) Separation of ascospores. An ascospore suspension was subjected to FACS. We determined that the population of cells centered at ~150 FSC-A and ~90 SSC-A (x1000) contained single cells whereas the other populations contained either aggregates or debris. (C) Separation of fluorescent and non-fluorescent ascospores. The target population from panel B was further separated into cells with low and high fluorescence (characterized in Table 3).
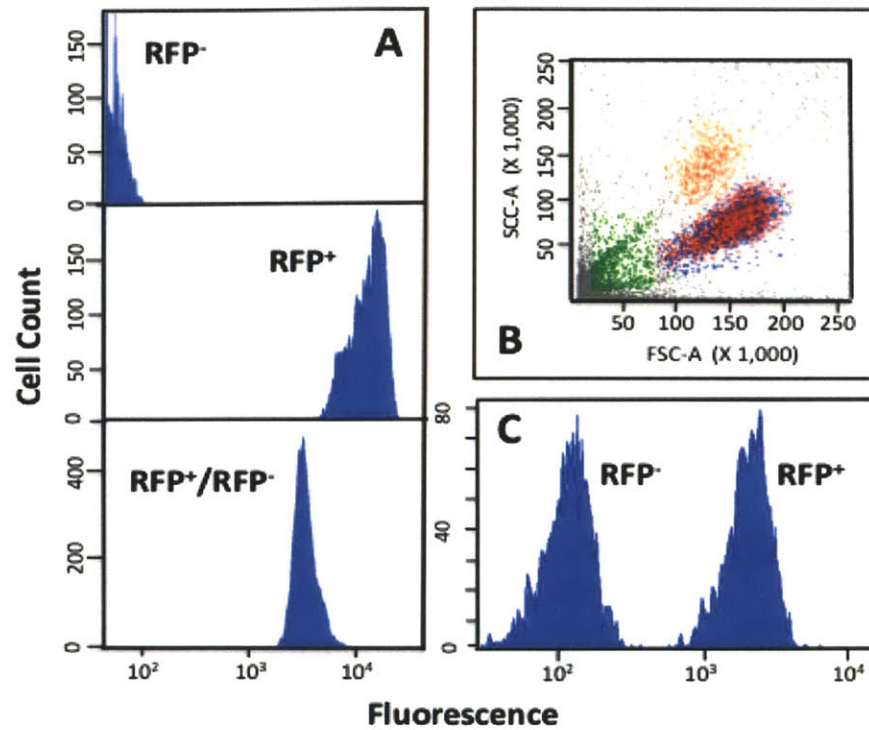
Figure A2-3 | **Growth characteristics of RFP[+] and RFP[-] Strains**. Strains ML1-4 (Table 1) were grown in a microtiter plate and the OD_{600} was recorded every 0.5 hr. OD's were converted to natural logs and the zero-time values were subtracted from each time point. YPD: yeast extract, peptone, glucose medium; SD: synthetic glucose medium supplemented for the requirements of the strains used (Amberg, Burke, and Strathern, 2005).
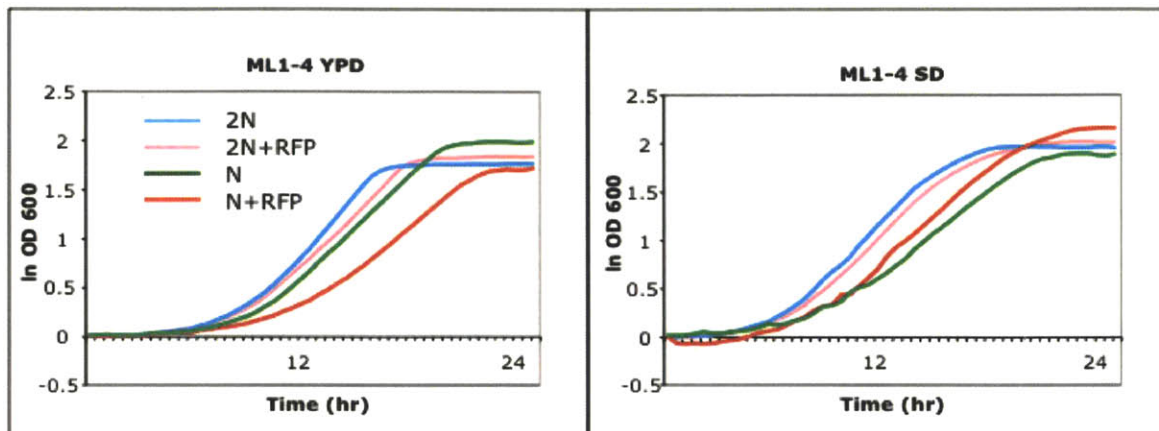
Table A2-1 | *S. cerevisiae* **strains used in the study.**

Strain	Genotype
ML1	*ura3-52/ura3-52 his3::hisG/HIS3 leu2::hisG/LEU2 trp1::hisG/TRP1 tec1::KANMX/TEC1 MATa (mata2::yEmRFP–HYGMX)/MATα*
ML2	*ura3-52/URA3 his3::hisG/his3::hisG leu2::hisG/LEU2 trp1::hisG/trp1::hisG tec1::KANMX/TEC1 MATa/MATα*
ML3	*ura3-52 leu2::hisG MATa (mata2::yEmRFP-HYGMX)*
ML4	*ura3-52 leu2::hisG MATα*
ML5	*his3::hisG trp1::hisG tec1::KANMX MATa*
ML6	*his3::hisG trp1::hisG tec1::KANMX MATα*

Table A2-1 | *S. cerevisiae* strains used in the study.

<u>Strain</u>	<u>Genotype</u>
ML1	*ura3-52/ura3-52 his3::hisG/HIS3 leu2::hisG/LEU2 trp1::hisG/TRP1 tec1::KANMX/TEC1 MATa (mata2::yEmRFP–HYGMX)/MATα*
ML2	*ura3-52/URA3 his3::hisG/his3::hisG leu2::hisG/LEU2 trp1::hisG/trp1::hisG tec1::KANMX/TEC1 MATa/MATα*
ML3	*ura3-52 leu2::hisG MATa (mata2::yEmRFP-HYGMX)*
ML4	*ura3-52 leu2::hisG MATα*
ML5	*his3::hisG trp1::hisG tec1::KANMX MATa*
ML6	*his3::hisG trp1::hisG tec1::KANMX MATα*

Table A2-2 | **Primers used in the study.**

Primer	Sequence
BCP538	5'-TGCAAACAACATCTCAACTCACTACTACCATTACTGTATTACTCAAAGAAGAAGCTTCGTACGCTGCA
BCP539	5'-TTTTTCTGTGTAAGTTGATAATTACTTCTATCGTTTTCTATGCTGCGCATATCGATGAATTCGAGCTCG
BCP569	5'-AGGCCTGTTAGAAAAGTGGAAAAACAAAT
BCP571	5'-AGGCCTTATCAGTTAGACCAATGTAATGAA

Table A2-3 | Characteristics of sorted ascospores.

Parameter Tested	Sorted Ascospores	
	RFP⁻	RFP⁺
Physical Count[1]	6.7×10^6	6.3×10^6
Viable Count[2]	4.2×10^6	4.5×10^6
Viability	63%	71%
Contamination with Hyg^R Cells[3]	0.16%	N/A
Contamination with Diploid Cells[4]	N/A	0.21%

[1]Counted in a haemocytometer.

[2]Serial dilutions were spread onto YPD plates and colonies were counted after 2 days at 30°.

[3]RFP-negative cells (2 X 10³ CFU/plate) were spread onto YPD plates containing 200 µg/ml of hygromycin-B. Colonies were counted after 3 days at 30°.

[4]RFP-positive cells (20-50 CFU/plate) were grown on YPD for 2 days at 30° and replica-plated onto lawns of a MATα tester strain. After 2 days at 30° the colonies were scored for halo formation. Diploids were implicated by lack of halo formation.

References

Amberg, D.C., Burke, D.J., and Strathern, J.N. 2005 Methods in yeast genetics: A Cold Spring Harbor Laboratory course manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA.

Churchill, G.A. and Doerge, R.W. 1994 Empirical threshold values for quantitative trait mapping. Genetics **138**: 963-971.

Conde, J. and Fink, G.R. 1976. A mutant of *Saccharomyces cerevisiae* defective for nuclear fusions. Proc. Natl. Acad. Sci. U.S.A. **73**: 3651-3655.

Darvasi, A. 1998 Experimental strategies for the genetic dissection of complex traits in animal models. Nature Genetics **18**: 19-24.

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

Dranginis, A.M. 1989 Regulation of *STA1* gene expression by *MAT* during the life cycle of *Saccharomyces cerevisiae*. Mol. Cell. Biol. **9**: 3992-3998.

Ehrenreich, I.M., Torabi, N., Jia, Y., Kent, J., Martis, S. et al. 2010 Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature **464**: 1039-1042.

Gianni, L., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., et al. 2009 Population genomics of domestic and wild yeasts. Nature **458**: 337-341.

Goldstein, A.L. and McCusker, J.H. 1999 Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. Yeast **15**, 1541-1553.

Guthrie, C. and Fink, G.R. 2004 Guide to yeast genetics and molecular biology. Meth. Enzymol. **194**: Abelson, J.N. and Simon, M.I., eds.

Jimenez-Gomez, J.M., Corwin, J.A., Joseph, B., Maloof, J.N., and Kliebenstein, D.J. 2011 Genomic analysis of QTLs and genes altering natural variation in stochastic noise. PLoS Genetics 7:e1002295.

Klar, A.J.S. 1980 Mating-type functions for meiosis and sporulation in yeast act through cytoplasm. Genetics **94**: 597-605.

Keppler-Ross, S., Nossz, C., Dean, N. 2008 A new purple fluorescent color marker for genetic studies in *Saccharomyces cerevisiae* and *Candida albicans*. Genetics **179**: 705-710.

Lander, E.S. and Schork, N.J. 1994 Genetic dissection of complex traits. Science **265**:2037-2048.

Steinmetz, L.M., Sinha, H., Richards, B.R., Spiegelman, J.I., Oefner, P.J., McCusker, J.H., and Davis, R.W. 2002 Dissecting the architecture of a quantitative trait locus in yeast. Nature **416**: 326-329.

Thacker, D., Lam, I., Knop, M., Keeney, S. 2011 Exploiting spore-autonomous fluorescent protein expression to quantify meiotic chromosome behavior in Saccharomyces cerevisiae. Genetics **189**: 423-439.

Timberlake, W.E., Frizzell, M.A., Richards, K.D., and Gardner, R.C. 2011 A new yeast genetic resource for analysis and breeding. Yeast **28**: 63-80.

Tong, A.H.Y. and Boone, C. 2007 High-throughput strain construction and systematic synthetic lethal screening in *Saccharomyces cerevisiae*. Meth. Microbiol. **36**: 369-386.