Leveraging natural isolates and experimental evolution to characterize biofilm-related phenotypes in *Saccharomyces cerevisiae*

Elyse A. Hope

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Maitreya J. Dunham, Chair

Joshua Akey

Aimée Dudley

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Leveraging natural isolates and experimental evolution to characterize biofilm-related phenotypes in *Saccharomyces cerevisiae*

Elyse A Hope

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

Biofilm formation is a protective community building behavior in which microbes participate to respond to external stress or colonize new ecological niches. It is a primary mechanism used by pathogenic yeast to persist on hospital surfaces and catheters and to invade the human body, and a useful trait in brewing and industrial fermentation. Despite the importance of understanding the phenotypic diversity and genetic basis of yeast biofilms, for reasons of experimental tractability they have typically only been examined in a subset of laboratory strains. The development of budding yeast *Saccharomyces cerevisiae* as model for biofilm development as well as the existence of growing collections of yeast natural isolates has created an ideal system for studying the spectrum of biofilm-related traits available to yeast and the primary adaptive routes by which they are acquired. We can observe biofilm-related

phenotypes experimentally using several assays including complex colony morphology, complex mat formation, flocculation in liquid media, agar invasion, and polystyrene adhesion. We have adapted these classic assays for use with an extensive collection of yeast natural isolates, and uncovered an enormous amount of diversity as well as new findings about the correlations between phenotypes and between phenotype and ecological and geographical niches. We also demonstrate that the previous understanding from lab strains of the effect of ploidy on these traits is more complicated in the context of these natural isolates. In further work, we focus on a single biofilm-related trait, flocculation, and its repeated evolution in continuous culture evolution experiments. We demonstrate, with genomic analysis of flocculent evolved clones, that there is one primary adaptive route to flocculation through changes in the regulation of known flocculin gene *FLO1*. Deletion of *FLO1* significantly reduces the rate of evolving the trait, and the evolution-based design methodology presented here represents a novel approach to evolutionary engineering.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Every step of the way, from my first research project in sixth grade, to this moment of receiving my PhD, I have been guided by incredible mentors. Thank you to Lisa Lockwood and David Bermudez, who started me on hypothesis-driven research, and to Bill Lamb and Rosa Hemphill who taught me that even young scientists can make novel contributions to their field and helped shape the research project that would be my driving force in high school and the basis of turning science research into a career. At Stanford, I was fortunate enough to be advised by Susan McConnell and Amy Ladd, who helped (and continue to help) this fledgling biologist figure out my passions and goals. When I joined Andrew Fire's lab as an undergraduate, his leadership style set the bar for the type of kind, patient mentorship that guides good, thorough science and a harmonious lab environment. I am grateful for this early demonstration of how wonderful working in a lab setting can be, and the experiences that guided me to look for a similar setting and mentorship style with Maitreya Dunham. Working with Maitreya as a trainee and as a colleague has been an absolute privilege; it is a rare thing, I realize, to reach the end of a doctoral degree and still like what I do and who I work for and with. My thesis committee members Josh Akey, Aimée Dudley, Ben Kerr, and Bob Waterston have been amazing resources, and I appreciate how they have challenged me to reach my potential and find my place in my field.

Also every step of the way, I have had the unerring support of my family and dear friends. Thank you to the remarkable female scientists in my cohort who have inspired me since undergraduate days (Cynthia He, Laura Brignone, Paras Fatemi, Christina Kratschmer) and throughout graduate school (Stephanie Battle, Rachel Gittelman, Alex Mason, Sonia Ting). I am fortunate to have a family who have always understood why I pursue this path. Thank you to all of you, especially to my Mom (Laura Hope) and my husband (Maxwell Libbrecht), for your loving encouragement and belief in me.

Chapter 1. INTRODUCTION

One of the earliest building blocks in our understanding of heredity was provided in 1866 by the monk Gregor Mendel and his famous pea plants. His study of traits including the color and shape variation of seed pods led to the development of the simple laws of inheritance that still hold up today for simple traits. These traits were binary (e.g. seed pods were either green or yellow, not a spectrum of colors in between) and appeared to occur independently from each other, two essential components to his work the importance (and rarity) of which Mendel could not have guessed. Initially, Mendel's laws were not well understood in the context of another school of thought, that of "blended inheritance", in which intermediate phenotypes were observed in a way that was not clearly explainable by Mendel's single gene, single trait results.

By the early 1900s, several experiments had demonstrated that some traits are controlled by the contributions of more than one gene, notably Herman Nilsson-Ehle's experiments with wheat coloration, published in German in 1909 (Nilsson-Ehle 1909) and summarized in an article by his contemporary, Wilhelm Johannsen, in The American Naturalist in 1911 (Johannsen 1911). In this article, Johannsen introduced the concepts and terms "genotype" and "phenotype", as well as fundamental ideas of modern genetics related to genotype-phenotype connections. These ideas included the concept that different genotypes might give rise to the same phenotype, that environmental factors can influence both, that seemingly homogenous populations might give rise to different phenotypes or a distribution of phenotypes, and that selection acts on standing genetic variation (although not yet in those terms). Importantly, in this period of thought, phenotype was coming to be thought of as something that could be influenced by

environmental factors while still retaining the same underlying genotype—our first look at one of the hallmarks of complex traits.

Most traits have phenotypes that vary along a spectrum, and these traits we would call complex, for exhibiting variation in the quantity of the trait produced, which can be continuous (e.g. weight) or non-continuous (e.g. numbers of bones), and influenced by environmental factors. If we can measure the trait in a quantitative way, a further distinction is made to label the trait "quantitative", although this distinction is rarely made precisely. An archetypal example of a quantitative trait is human height, which is not a binary trait inherited with clear rules of dominance, and can be influenced by environmental factors like nutrition or smoking. The genetic basis of human height has been a difficult problem to solve, with decades of work by hundreds of investigators ultimately yielding associations between height and more than 400 genomic regions, which still only explain up to 60% of the variation in this complex trait (Wood et al. 2014).

Many human diseases have actually been revealed to be "Mendelian", or caused by mutations in single genes that are inherited according to Mendel's laws of assortment, segregation, and dominance. One well-studied example of a disorder that is considered Mendelian is cystic fibrosis, an autosomal recessive disease. CF is caused by loss-of-function mutations in the *CFTR* gene, which encodes a protein that regulates ion transport across the inner membrane of tubes in the body, including sweat ducts and the lungs. Loss of transporter function leads to changes in the content of fluids at that membrane, ultimately facilitating infection in the lungs by pathogenic bacteria, most effectively by *Pseudomonas aeruginosa* (Wang et al. 2014; Rowe, Miller, and Sorscher 2005). Even in this genetically well-understood disease, 2,009 different variants of CFTR have been associated with the disease, some common and some

extremely rare (http://www.genet.sickkids.on.ca/cftr/app), and despite its label as a Mendelian trait there is variation in how the CF phenotype manifests, sometimes influenced by environment or other genetic factors (Bombieri, Seia, and Castellani 2015). Moving away from Mendelian diseases and into complex diseases means untangling some of the greatest medical challenges of our time, including cancer, heart disease, diabetes, and Alzheimer's, which have no major effect causal disease gene.

Although human disease is a useful vehicle for understanding complex traits, the majority of important traits across all organisms are complex. On the front lines of agricultural research, geneticists are working to better understand complex traits like drought tolerance, pest resistance, and root depth (Samayoa et al. 2015; Lou et al. 2015; J. Zhou et al. 2007) that will assist in intelligent breeding and genetic engineering decisions in response to climate change. In the microscopic world, bacteria and other microbes exhibit a variety of complex traits that enable them to colonize new nutrient sources and resist harmful chemical treatments, contributing to pathogenicity and the development of drug resistance (e.g. Ehrenreich et al. 2010; Anderson et al. 2011). The motivation for seeking a better understanding of the genetic underpinnings of such complex traits is clear, but new techniques are needed to identify the many contributing genes and their roles in the generation of a complex phenotype. Complex trait genetics is one of the most important areas of study in modern genetics and genomics.

1.1 Quantitative trait mapping methods: principles, strengths, and shortcomings

Every approach to map complex traits is constrained by the incredible variety of ways in which they can occur. A trait can be considered complex if multiple genotypes cause the same phenotype, for example mutations in genes in a pathway might share a phenotype, or multiple

different combinations of alleles could ultimately yield the same phenotype as well (Steinmetz et al. 2002). A single gene can also affect multiple phenotypes (pleiotropy), and not always in the same way; it is rarely accurate to describe a single gene/single trait relationship (reviewed in Mackay, Stone, and Ayroles 2009). Genotype is also associated not with the guaranteed occurrence of a trait but rather with the probability of the occurrence of a trait, as in the case of "incomplete penetrance" where the causal allele is present but the phenotype is not always observed due to the influence of environmental factors or other genes (reviewed in Lander and Schork 1994). Such cases confound mapping studies in which test and control groups are created based on phenotype; a causal allele could show up in both populations in a case of incomplete penetrance. Environmental factors can also cause multiple phenotypes to result from the same genotype (Lander and Schork 1994) or cause a phenotype to occur with no genetic basis at all. These "context dependent effects" are an important additional dimension of complex traits, and can include environmental effects, sex, and epistasis (Mackay, Stone, and Ayroles 2009; Dilda and Mackay 2002; Flint and Mackay 2009; Stam and Laurie 1996). Despite the intricacy of complex traits, identifying alleles with a large effect on complex traits has been reasonably successful using a variety of approaches.

In model organism *Saccharomyces cerevisiae,* many of the candidate genes potentially contributing to complex traits have been identified agnostically, with random mutagenesis followed by screening to look for an effect on a phenotype of interest (e.g. Mösch and Fink 1997) or systematic deletion of all genes individually (Shoemaker et al. 1996; Giaever et al. 2002; Winzeler et al. 1999). One advantage of this approach has been that once a mutation with an effect is identified, it can be easily isolated and mapped, e.g. with barcode sequencing for modern deletion collections (Shoemaker et al. 1996); however, such collections and screens are

laborious to generate in multiple backgrounds. In more complex organisms like mice and flies, this approach works poorly for traits with a wide range of quantitative variation or that are not observed due to incomplete penetrance. It can also present a misleading picture with respect to polygenic traits, making causality appear simpler than it is (reviewed in Flint and Mackay 2009). To generate a more complete picture of the genetic basis of both Mendelian and complex traits, there are two primary mapping approaches, each also with strengths and shortcomings. The first approach, association testing, asks if a trait and an allele are associated in a population; on a large scale these are called Genome Wide Association studies (GWAS). The second approach, linkage mapping, asks if a trait and an allele are inherited together and necessitates relatedness, which is not a component of association mapping (Lander and Schork 1994; Mackay, Stone, and Ayroles 2009). Linkage mapping is commonly referred to as Quantitative Trait Loci (QTL) mapping, though both linkage and association testing are used to map QTLs (the genetic loci underlying complex traits). For studies of human traits, both approaches are highly complicated by the heterogeneity, comparatively short existence of modern species, and long generation times of humans. Model organisms possess many features that make them ideal tools for the study of complex traits, including short generation times, smaller genomes, prodigious reproduction, high meiotic recombination rate, and the feasibility of both inbred and outcrossed lines (reviewed in Liti and Louis 2012).

1.1.1 *Association mapping is complicated by population heterogeneity*

The premise of association mapping is to search for alleles that occur at a statistically significant higher frequency in a population with a trait of interest compared to a control population lacking that trait (Lander and Schork 1994). This approach uses linkage disequilibrium between known markers and nearby causal variants to localize genomic regions associated with the trait of

interest and depends on historical recombination rates estimated for a species to determine how large haplotype blocks are likely to be. The most common use of association mapping is currently GWAS, which search for associations typically with a disease trait across the whole genome for large affected and control populations. GWAS have made notable progress identifying common variants underlying Crohn's disease, type 2 diabetes, and others (reviewed in Donnelly 2008), and studies are ongoing. One strength of GWAS is that they can reveal instances of pleiotropy and connections between traits, if the same loci end up associated with multiple traits and associations with other phenotypes like ethnicity are ruled out. The roles of genes associated with disease traits can provide important insight into the mechanisms of disease that would be harder to detect without this type of work (Donnelly 2008), especially insights into overlapping mechanisms across diseases (e.g. the role of interleukin receptors in both Type 2 Diabetes and Multiple Sclerosis) (Frazer et al. 2009). Despite these successes, GWAS have fallen short of expectations, revealing only a small part of the genetic basis especially of human disease traits, a problem frequently termed "missing heritability." Variants associated with disease are typically common, because they have to be prevalent enough in the population for a significant association to be detected; one of the potential sources of missing heritability is a large proportion of causal variants too rare in the population to be detected in a GWAS. Additionally, causal variants could exist with phenotypes too weak to detect by this approach (Bloom et al. 2013). Even in model organisms, GWAS are limited; insufficient mosaicism between genomes of strains that might be sampled by this method makes it hard to detect more than a few associations, although this can potentially be resolved by intentional outcrossing (Liti and Louis 2012). Missing heritability has been most successfully investigated in model organisms using linkage mapping approaches instead (Bloom et al. 2013). Estimates of missing

heritability themselves might also be limited by some of the possibly false assumptions underlying their calculation, since common methods of estimating heritability do not account for epistasis and genetic interactions between loci (modeled in Zuk et al. 2011).

Another issue in association studies is the potential association between an allele and a trait not because of causality but because of an unseen relationship between both of them and another factor. An example of this issue is the discredited relationship between an allele of dopamine receptor *DRD2* and alcoholism (Blum et al. 1990; Gelernter, Goldman, and Risch 1993; Cloninger 1991; Lander and Schork 1994), which was confounded because of additional variation between the control and test groups. Ideally, the control and test groups would be from the same genetic background and exposed to the same conditions, so the only differences between them would be related to the specific phenotype of interest; however, in many human studies it is simply not possible to control for these variables between groups of test subjects.

Further complicating this picture, it is difficult to separate causal alleles from other alleles close by in the genome; an allele could have a positive association with a disease because it is itself causal, but it could also simply be linked to the causal allele due to the short genetic distance between them reducing the frequency of recombination (Lander and Schork 1994). Even in the era of (relatively) low-cost whole genome sequencing, human genetic data sets are still more likely to be SNP genotyping of several markers or exome sequencing, instead of whole genome sequencing; for mapping studies this makes it especially difficult to resolve the causal variant. Some of these challenges from association mapping studies are resolved, especially in model organisms, by using linkage mapping.

1.1.2 *Linkage mapping is complicated by small effect QTLs*

Differences in the background variation between comparison groups have also contributed to discrepancies in predicted QTLs and effect sizes in linkage mapping studies (reviewed in Flint and Mackay 2009), but for identifying and resolving QTLs in model organisms, linkage mapping has largely proven a very successful technique. One of the earliest examples of linkage mapping was put forward by Karl Sax in 1923, with patterning and size differences in beans (Sax 1923). By the time of this work, it was commonly accepted that while some traits are Mendelian or, as Sax termed them, "single-factor", many more are "multi-factor" traits with contributions from many genetic loci. Sax tracked multiple seed phenotypes through bean crosses based on segregation patterns in progeny tried to determine whether each is a single- or multi-factor trait and how many factors were involved.

Later work enabled localization of these "factors" to actual genomic regions, and the premise of subsequent linkage mapping studies has been to correlate inheritance patterns of a trait with inheritance patterns of a chromosomal region or regions (Lander and Schork 1994). Early linkage mapping depended on the existence of detailed genetic maps of "markers", typically known restriction digest patterns that facilitated comparison of chromosomal fragments. These markers, called Restriction Fragment Length Polymorphisms, are created by digesting the genome with restriction enzymes and hybridizing the resulting fragments to DNA probes; the size of the fragments and whether or not they hybridize determines if there is a mutation in that region compared to an expected reference (Botstein et al. 1980, adapted for human genetic mapping). Comparative inheritance of these markers and traits of interest facilitated the mapping of the genetic regions underlying both Mendelian and complex traits. This technology has largely been replaced by whole genome sequencing for model organisms

but the core principles of this type of linkage mapping apply to these more precise and higher resolution marker sets.

Early linkage-based QTL mapping suffered from high numbers of false positives due to the numbers of loci being evaluated independently (e.g. the mean phenotype for strains with marker AA was tested against the mean phenotype for strains with marker AB and a null hypothesis of no phenotypic difference, for every possible marker), and challenges in detecting the difference between small effect and distant linkage. In 1989 Lander and Botstein revolutionized the field of QTL mapping with their "interval mapping" technique, which uses LOD scores, a calculation of linkage that in this application can estimate how likely it is that the data are explained by the presence of a QTL, to facilitate the analysis of many markers at once rather than each independently. Interval mapping maximizes a likelihood function using genotypes described by a probability distribution based on distance (and therefore probability of recombination), rather than the known genotype at a given locus. Lander and Botstein defined a LOD score threshold above which a QTL can be declared for a given genome size and marker density. One advantage of this approach is that it generates a distribution of LOD scores across the whole genome, eliminating the false discovery threshold issues from multiple hypothesis testing; another advantage is that it provides interval information around putative QTLs (Lander and Botstein 1989).

This technique of interval mapping was first used by Paterson *et al* (in a companion paper to Lander and Botstein) to map three continuously varying traits in tomatoes: weight, acidity, and concentration of soluble solids (primarily sugars). They identified several strong effect QTLs for each trait (between four and six) but also acknowledged the limitations of the study; they were unable to resolve the individual contributions of linked genes or of smaller effect QTLs, and they

identified only one QTL previously associated with soluble solids concentration in a study that employed selection and backcrossing, a classic technique to increase the amount of variation explained by those QTLs by enriching for causal loci (Paterson et al. 1988).

Modern linkage mapping studies have been assisted by the information provided by whole genome sequencing. Especially for model organisms with smaller genomes, it is common to have available WGS information rather than SNP genotyping, and the natural variation in model organisms revealed by whole genome sequencing data (e.g. 0.57% average pairwise divergence across the genome for 32 *S. cerevisiae* natural isolates (Liti et al. 2009)) makes them ideal for linkage mapping studies. Modern mapping approaches have also effectively incorporated other new data types, including genome-wide gene expression data. Gene expression can be treated as a quantitative trait in its own right since it shows variation among segregants in a cross, and when combined with molecular marker information like genome sequence it can be used for expression QTL or eQTL mapping. This technique identifies loci linked to the expression of a gene of interest (Jansen and Nap 2001), and has been used to query some of the more intricate and previously less accessible aspects of complex trait genetics, including pleiotropy, epistasis, and the organization of genes into pathways. Sometimes the linkage between expression and an eQTL will be in *cis*; in eQTL mapping from a cross between *Saccharomyces cerevisiae* strains BY4716 (a laboratory strain derived from S288C) and RM11-1a (derived from California vineyard isolate Bb32(3), collected by Robert Mortimer), Brem *et al* estimated that 36% of the linked transcript variation was *cis*, linked to its own gene or regulatory region. Other associations will be in *trans* (association between transcript and other genes) which are informative for constructing putative pathways (Mackay, Stone, and Ayroles 2009; Jansen and Nap 2001). Brem *et al* identified eight groupings of genes with shared linkage function in

their BYxRM yeast cross, each associated with a defined role in the cell, including mating, leucine and uracil biosynthesis, and budding (Brem et al. 2002). The association between genes in these linkage groups and expression of a putative regulator (or indirect regulator in the case of metabolic enzymes) provided useful data for inferring mechanistic information. Additional characterization of the BYxRM cross by Yvert *et al* clustered the gene expression data and found 76% of their clusters contained genes on different chromosomes (eliminating proximity as the driver of clustering). eQTL mapping using these clusters indicated that 80% of the linked QTLs were *trans* (Yvert et al. 2003), a higher estimate than in Brem et al. It is also possible to identify eQTL hotspots, genomic regions associated with increased expression of many transcripts (reviewed in Mackay, Stone, and Ayroles 2009), a method for detecting pleiotropy. Although eQTL studies can provide direct mechanistic insights, like in early association mapping studies the high number of markers and transcripts but typically low sample size makes them prone to difficulties with multiple hypothesis testing and high false discovery rates (Mackay, Stone, and Ayroles 2009).

The biggest challenge for all QTL mapping techniques is still improving resolution of causal alleles among high numbers of both large and small effect QTLs. Current mapping techniques often identify genomic regions with many genes, only one of which may be causal, or several of which may be small effect QTLs; because the close genetic distance in these regions makes recombination rare (Steinmetz et al. 2002), it is challenging to identify a causal allele or, even more desirably, a causal nucleotide/set of nucleotides. A number of new efforts using multiple mapping approaches in combination have improved the detection of small effect QTLs and the resolution of casual Quantitative Trait Nucleotides (QTNs).

1.1.3 *Using combination approaches to identify the contributions of small effect QTLs*

Small effect QTLs are loci for which the average difference in the trait across the two genotypic classes is small when normalized by the standard deviation in the trait, e.g. for a trait with a small or highly variable range of quantitative values (Mackay, Stone, and Ayroles 2009). Detecting their linkage to a trait remains an ongoing challenge in QTL mapping, especially for human traits; higher numbers of progeny can boost this signal, but are a luxury of model organism research. Ongoing efforts to resolve small effect QTLs are motivated by their potential contribution to "missing heritability", and their known role in complex traits; indeed, early studies that detected moderate numbers of QTLs of large effect underlying complex traits turned out to be underpowered, and subsequent work has revealed that most complex traits have more QTLs of small effect (Mackay, Stone, and Ayroles 2009), with an exponential distribution of allelic effects (few alleles of large effect and increasing numbers with small effect) (Flint and Mackay 2009). The existence of this issue was predicted by Lander and Botstein in 1989 who modeled, after a formula by Wright (Wright 1968) the relationship between the number of QTLs segregating in a cross as a function of phenotypic difference between the two starting strains, and demonstrated that a small number of QTLs would be detected only for a cross segregating large effect sizes. To address this issue, Lander and Botstein also pioneered the idea of reducing the number of progeny needed to map a trait by genotyping only the most phenotypically extreme progeny, typically the extreme 5%. This concept was taken to its utmost two decades later with an "X-QTL" or extreme QTL study aimed at identifying QTLs underlying drug resistance traits in yeast; the study generated unusually large populations segregating the traits, under selection to enrich for extreme phenotypes for further analysis (Ehrenreich et al. 2010).

This type of mapping also exemplifies how the most effective studies analyzing QTLs use a combination of different approaches. The first method employed in X-QTL mapping is the generation of high numbers of segregants (~10^7) in a cross with extreme phenotypes driven by selection; in the 2010 study from Ehrenreich *et al*, these segregants had chemical resistance. The authors evaluated chemical resistance phenotypes of the progeny across a gradient of conditions and applied a Bulk Segregant Mapping strategy to pool the most phenotypically extreme progeny with and without the trait being evaluated, and look for genomic variants linked to those traits. For a sample quantitative trait, sensitivity to DNA damaging agent 4-nitroquinolone, conventional QTL mapping had identified one major effect QTL, and with additional backcrossing another minor effect QTL. X-QTL mapping and subsequent validation revealed nine additional QTLs contributing significantly to this trait (Ehrenreich et al. 2010).

Selection can also be used to improve the resolution of small effect QTLs by using complex outbred populations under selective pressure to maintain the loci connected to the trait of interest and reduce the likelihood of fixing variation from random genetic drift (reviewed in Flint and Mott 2001). Other studies in model organisms have also had success eliminating large effect QTLs through crosses and leaving the smaller effect QTLs to identify. For example, following up on a study that revealed three strong effect alleles with complex interactions underlying high temperature growth in yeast (Steinmetz et al. 2002), Sinha *et al* performed additional backcrossing and selected segregants with high temperature growth phenotypes but lacking the strong effect alleles (Sinha et al. 2008). This approach revealed three additional small effect QTLs linked to the trait, one of which was also physically close to a major effect QTL. Further analysis revealed that these linked small effect and major effect QTLs had a background-

dependent epistatic interaction, demonstrating how complex the interactions between QTLs can be and the importance of continuing to work to uncover small effect QTLs.

Although identifying small effect QTLs can add significant mechanistic insights to the biological basis of complex traits, from a practical perspective they also add more loci to test experimentally. To identify specific causal alleles or QTNs while including small effect loci for consideration, the most successful studies have employed combinations of different mapping approaches and incorporated functional validation techniques.

1.1.4 *Using combination approaches to identify causal allelic variants*

Earlier work with mutagenesis and deletion collection studies has identified candidate genes with known involvement in many complex traits of interest. Although the overlap between genes identified as important in mutagenesis studies and mapping studies is highly variable (Dilda and Mackay 2002), candidate genes within QTL peaks provide a natural starting point for validation. Functional complementation testing, in which the wild-type allele is added through a cross or on a plasmid to see if the wild-type phenotype is restored, can be performed for each candidate gene under a QTL peak individually to look for a change in phenotype (Flint and Mott 2001), and allele replacements of candidate genes can perform a similar function. Another useful modern method for identifying causal alleles has been to look directly at protein abundance data for genes within putative QTLs. In an early experiment of this type, Stam and Laurie (Stam and Laurie 1996) characterized the impact of individual polymorphic sites on alcohol dehydrogenase expression in fruit flies. Approaches using higher numbers of segregants and pooling them by phenotype have also been useful to reduce the impact of variation between experiments and the prevalence of unlinked background variants (Jansen and Nap 2001). This Bulk Segregant Analysis method can also improve sensitivity to small effect QTLs (Liti and Louis 2012). Most

studies that ultimately identified QTNs down to the level of exact causal polymorphisms have employed a combination of mapping and functional genomic approaches.

One classic study in yeast (Steinmetz et al. 2002) examined the genetic basis of a high temperature growth phenotype in a cross between laboratory strain S288C and a clinical isolate, an important phenotype to understand since growth at higher temperatures can contribute to pathogenic yeast's ability to survive and reproduce inside the human body (Steinmetz et al. 2002; Liti and Louis 2012; Sinha et al. 2008). The authors employed marker genotyping with segregants to identify QTLs linked to the trait followed by comparative genomics approach to identify which of these the alleles were most likely to contribute to high temperature tolerance. By comparing the genomes of several high temperature clinical strains and low temperature strains across the QTLs identified in their cross, the authors identified candidate genes specific to this study based on shared features of the high temperature natural isolates. They also used gene expression analysis to validate which of the candidate genes had increased expression associated with the trait, and their results were inconclusive, with no significant differences in one of their major QTLs. This finding was consistent with this locus being a "transgressive QTL", with a phenotype antagonistic to the expected phenotype based on the parent strain of origin (Liti and Louis 2012). Even though the laboratory strain was selected for their study based on its low temperature phenotype, an allele from that strain actually causes a high temperature phenotype not observed in the context of that strain background. This complicated their mapping efforts and led to the development of an important new technique for identifying QTNs in the context of other variants.

In reciprocal hemizygosity analysis, the performance of each candidate allele is examined individually in the context of an isogenic hybrid background by deleting the other copy of the

allele. This technique is made even easier in yeast by the existence of deletion collections (reviewed in Liti and Louis 2012). The resulting strain is diploid except at the query location, and it is possible to directly test the contribution of each allele to the trait of interest. An advantage of this method over allele swaps in the parent strains is that having the background variants present from both parent strains permits combinatorial effects. For example, if the allele only drives a particular phenotype in the presence of alleles from its own background that would be observed; but instances of transgressive alleles that drive a phenotype different than expected in the context of the alternate background can often be detected as well. Like allele swaps in the parent strains, this method is agnostic to which strain is contributing the phenotype, but it also makes it possible to resolve more complex genetic relationships involving variants from both backgrounds. Reciprocal hemizygosity also more easily facilitates queries of essential genes, since the hybrid will always have at least one copy. Using this technique, Steinmetz *et al* were ultimately able to systematically test the effect of every allele within the QTLs they identified at multiple temperature conditions and identified three genes associated with the high temperature growth trait, one contributed from the low temperature parent (Steinmetz et al. 2002).

Reciprocal hemizygosity is a useful tool for functional genomics validation of candidate genes, and for testing hypotheses particularly pertaining to large effect loci (Lee, Magwene, and Brem 2011). Lee *et al* demonstrated the efficacy of this approach using reciprocal hemizygotes for alleles of gene *CDC28* from two natural isolates of yeast, one from an Ivory Coast ginger beer strain and the other from a North American oak strain. *CDC28* was a known candidate gene for differences in cell morphology from previous work and was linked to an unusual meiotic spore arrangement in progeny from this study; it also had a unique polymorphism in the Ivory Coast strain, which also uniquely harbored the unusual spore arrangement observed in the

progeny. The phenotype conferred by the Ivory Coast allele of *CDC28* was stronger in the Ivory Coast background, indicating other mutations also modify the trait, e.g. small effect QTLs that could likely be identified by additional approaches. Not only did reciprocal hemizygote analysis confirm the role of *CDC28* in these traits, it also demonstrated the power of this particular type of functional validation: although it could not be used to identify the other loci contributing to the trait, it did reveal that they existed.

In both of these studies employing reciprocal hemizygosity, the use of natural isolates played an essential role in their success at identifying causal alleles. The comparative genomics approach used by Steinmetz *et al* to identify promising candidates relied on having multiple clinically-derived natural isolates with a high temperature growth phenotype to compare, and Lee *et al* used comparison with other natural isolates to determine that the spore arrangement observed in their Ivory Coast strain was unique among the natural isolates and that they should be looking at unique genetic variants as well.

In another combined approach study, natural isolates were used to resolve QTLs to their actual causal QTNs. Gerke *et al* explicitly rejected the use of laboratory strains for mapping studies, since the many deletions and auxotrophies that are designed to make them more functional in the lab could obscure the effects of natural variation (Gerke, Lorenz, and Cohen 2009). They instead focused on a trait known for its high variability in natural isolates, sporulation, which was also hypothesized to be strongly influenced by the selective pressures on natural isolates in different environmental niches. Using shotgun sequencing and QTL mapping they identified five loci significantly linked to sporulation efficiency, four from the high efficiency parent and one transgressive QTL from the low efficiency parent; three of these were major effect QTLs. Two of the major QTL peaks already had a candidate gene in the region,

which validated with reciprocal hemizygosity; the third major QTL revealed a candidate gene after multiple genes were evaluated using reciprocal hemizygosity. For one gene, the parental strains diverged only in the regulatory variation and the QTNs were easy to identify; for the other two, the causal polymorphisms were identified with systematic allele replacement of divergent sites. They recorded additive interactions between the QTNs, indicative of a detection of epistatic interactions. Genome analysis of other natural isolates from similar ecological niches as the parental strains in this study revealed the abundance or rarity of each of these alleles, providing a framework for understanding which might be most important for adaptation to a given ecological niche. This work took complex trait analysis down to the level of causative polymorphisms and placed the work in a biological context. It represented the cumulative power of each of the approaches used together in a single study, and the contributions natural isolates can make to mapping complex traits, both as a source of comparative genomic information and of phenotypic and genetic diversity.

1.1.5 *Experimental evolution provides an alternative avenue for studying complex traits*

Across studies that take a combined approach there are key uniting features. One of these is the use of candidate genes, previously identified in mutagenesis or deletion collection experiments or through gene expression data, to narrow down QTLs to likely causal alleles. Another hallmark of these studies is the use of selection to enrich for phenotypically extreme progeny and increased effect size of QTLs. Both of these features connect mapping studies to another technique that can provide insight into complex trait genetics: experimental evolution.

In evolution experiments, microorganisms (typically bacteria or yeast) are propagated under selective pressures to examine how the genome responds. Such experiments have provided insight into the types of adaptations most common during experimental evolution, including

aneuploidy, copy number variation, transposable element insertions, and other complex genome rearrangements. It is also an important tool for identifying the genes that contribute most frequently to the evolution of complex traits, providing a method not only for identifying candidate genes for mapping studies but also a potential approach for identifying adaptive alleles that would be of small effect in QTL studies and contributing alleles not found in natural isolates. This unbiased approach is useful not only for identifying contributors to complex traits that might be masked in mapping studies but for identifying which alleles are the most useful for adaptation.

Evolution experiments are commonly performed using a limited nutrient supply. Hong and Gresham (Hong and Gresham 2014) evaluated the adaptive strategies of yeast evolving with a limited supply of nitrogen. This experiment was conducted using chemostats, continuous culture devices that maintain a constant growth rate and nutrient limitation compared to the fluctuating nutrient levels of serial dilution techniques. The mechanism of and genes involved in nitrogen uptake are well understood in yeast, providing a biological framework for their results. Their experiments revealed multiple common adaptive routes, including increasing the copy number of nitrogen transporter genes, remodeling nutrient response signaling pathways, and co-evolution of a group of alleles with epistatic relationships. These results demonstrated how adaptation to nitrogen limitation is similar to other nutrient-limited systems (transporter gene amplifications being a mechanism shared with glucose- and sulfate-limited experiments) (Brown, Todd, and Rosenzweig 1998; Gresham et al. 2008) and also how experimental evolution can identify important adaptive genetic interactions that would be harder to detect with mapping techniques.

Like in QTL mapping approaches, the use of natural isolates is also providing promising advances in experimental evolution. Yeast employed in industrial applications is often exposed to conditions inhospitable for the typical laboratory strain; instead of adapting a laboratory strain to a new environment, working with natural isolates creates an opportunity to sample natural diversity for strains already well-suited to that environment and evolving that strain from a stronger initial starting point (reviewed in Steensels and Verstrepen 2014). Examining adaptation in the context of different genetic backgrounds provides access to phenotypes not available from working with laboratory strains and even differences in the favored types of adaptation, e.g. from differences in transposable element burden (Liti et al. 2005).

1.1.6 *The power of genetic diversity in complex trait analysis*

Despite the progress made in mapping complex traits using the full history of experimental and analytical methods, all of these approaches are limited by known phenotypes and, for linkage mapping, the genetic diversity available in single crosses. The most successful linkage mapping studies have been able to move beyond what we know from laboratory strains and leverage natural isolates to access new phenotypes and identify candidate genes using comparative genomics.

One of the early essential ideas in QTL mapping, as presented by Lander and Botstein in 1989, is that of identifying crosses with sufficiently large effect sizes that are likely to be promising and effective for QTL mapping. Natural isolates provide a system to sample more possible effect sizes than might be accessible from continued work in laboratory strains alone, and as such are frequently used in QTL mapping crosses; however, such studies are still limited by their pairwise nature. QTL mapping is limited to the genetic and phenotypic variation being sampled in each particular cross; there is no guarantee that the loci identified approach the

number truly involved in the phenotype or that they are the most commonly represented causal loci in the population (reviewed in Flint and Mott 2001). Work with expanded collections of yeast natural isolates helps ensure that we are detecting the range of phenotypes actually available, allows us to determine the most successful adaptive routes for evolving those phenotypes in competitive environments, helps us understand the subfunctionalization of pleiotropic genes (Flint and Mackay 2009), and provides opportunities to identify QTLs that can only be detected or only have their effect in certain backgrounds (reviewed in Liti and Louis 2012), to understand shared features of those backgrounds, and to describe the role of environmental factors on the genotype/phenotype connection.

1.2 NATURAL ISOLATES PROVIDE A WEALTH OF GENETIC AND PHENOTYPIC DIVERSITY FOR STUDYING COMPLEX TRAITS

1.2.1 *Natural isolates are the exceptions to the rules from laboratory strains*

The study of yeast natural isolates has already challenged some of the tenets of yeast biology developed from work with domesticated laboratory strains. Contrary to the expectations of previous work (Barnett, Payne, and Yarrow 1983), Warringer *et al* identified significant phenotypic differences between *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* when a broad panel of phenotypic data were collected across a collection of natural isolates (Warringer et al. 2011). Generalizations developed from working with laboratory strains regarding basic concepts like ploidy are also being cast into doubt as close relationships are revealed between ploidy, phenotype, and environmental factors in natural isolates (Zörgö et al. 2013). Another notable example of natural isolates improving our understanding of phenotypic variation in yeast is that of our growing understanding of prions. Prions, a term for proteins that exhibit self-

templating, heritable changes in protein conformation, were previously thought to be a type of yeast "disease" and considered a nuisance to laboratory experiments. Recent work from Halfmann *et al* demonstrated that, far from being a phenomenon exclusive to laboratory yeast, prions are found at low frequencies among natural isolates, and are now thought to be a useful mechanism for fast adaptation to environmental changes (Halfmann et al. 2012). Prion [PSI+], found in 10 of the 690 natural isolates tested in their study, is a misfolded version of the protein Sup35, which when functioning normally facilitates translation termination. When sequestered into its prion form, Sup35 is unable to perform its usual function, and increased incidence of stop codon read-through provides the yeast access to cryptic variation in the 3'UTR that can cause dramatic shifts in phenotype that would have taken generations to access with the normal rate of mutation. In the event that some of these phenotypes prove advantageous to the wild yeast in its ever-changing environment, there is a background rate of persistence (~1 in 10^6 cells for [PSI+]) of the prion form of proteins in yeast natural isolates to ensure continued exploration of this potentially advantageous adaptive space. Indeed, Halfmann *et al* found that growth advantages under different conditions, including drug treatments, were lost among strains cured of prions. That the study of natural isolates facilitates a greater understanding of how yeast adapt to new environments is a compelling reason to look outside the lab.

1.2.2 *Natural history of yeast revealed: phylogenies and phenotypes*

The genetic diversity supplied by yeast natural isolates has also answered questions about speciation and hybridization in yeasts, with broad applications to other organisms. In 2006, Liti *et al* (Liti, Barton, and Louis 2006) introduced 13 wild isolates of *S. cerevisiae* that would later become part of the ongoing *Saccharomyces* Genome Resequencing Project, though sequenced first only at the level of single genes. They were able to utilize crosses between these and natural

isolates from other *sensu stricto* species to describe a correlation between increased sequence divergence and decreased gamete viability, and build rough phylogenies. More recent work with natural isolates has further explored phylogeny of wild yeasts, their genetic relatedness across ecological niches, the role of human activity in their history, and their broad range of possible phenotypes.

Although natural isolates have been collected for decades, the early 2000s brought renewed interest in studying the phenotypic and genotypic diversity they provide with modern genomic tools. Small collections of natural isolates yielded insight into the phenotypic variation (Fay et al. 2004), genetic variation (Doniger et al. 2008; Fay and Benavides 2005) and selective pressures on natural isolates (Aa et al. 2006), and a large collection of 651 strains presented by Legras *et al* in 2007 revealed, from microsatellite genotyping, that almost 30% of the variation could be accounted for by geographical distance (Legras et al. 2007). By 2009, Schacherer *et al* presented SNP genotyping of 63 natural isolates (Schacherer et al. 2009) and Liti *et al* presented an expanded version of their collection with associated low-coverage whole genome sequencing data (Liti et al. 2009). Though advances in sequencing methods have since improved our understanding of these strains, and this strain collection has itself improved (Cubillos, Louis, and Liti 2009), this initial paper was one of the first glimpses into the relatedness and population history of a large collection of wild yeasts using whole genome information, including quantities like pairwise sequence divergence and estimates of linkage disequilibrium. Notably, their comparison between *Saccharomyces paradoxus*, which has had little involvement in human activities, and *Saccharomyces cerevisiae*, a workhorse for fermentation throughout human history, revealed corresponding amounts of population admixture, with *S. paradoxus* strains grouping into clean lineages along geographic boundaries and many *S. cerevisiae* strains

showing significant mosaicism with contributions from the North American and European or wine strains, with the greatest corresponding human influence. Subsequent work with larger collections of yeasts has refined this view of *S. cerevisiae* population structure to be predominantly shaped by geographic origins, with some fine-scale variation linked to ecological niche, with the exception of heavily admixed fermentation strains (Cromie et al. 2013).

A subset of the Liti strains were included in the extensive phenotypic characterization of natural isolates performed by Warringer *et al* (Warringer et al. 2011) and later Skelly *et al* (Skelly et al. 2013), generating a detailed dataset of gene expression and phenotypic data across natural isolates for *S. cerevisiae* and other species. The 200 conditions under which Warringer *et al* collected growth curves for the natural isolates represented the types of selective pressures (including changes in carbon source, nutrient availability, exposure to toxins, and more) yeast are more likely to experience outside of the lab. Importantly, including laboratory strain S288C in their phenotypic panels demonstrated that the laboratory strain is actually a phenotypic outlier, clustering far from other strains of *S. cerevisiae* across metrics including carbon utilization, toxins, and nutrient requirements. They recognize explicitly that using such a phenotypically divergent laboratory strain as the gold standard for behavior of *S. cerevisiae* is troubling. Going even more in depth into the phenotypes of natural isolates, Skelly *et al* created a set of "phenomics" data across more than 14,000 characteristics at steady state growth conditions for 22 isolates from the Liti collection, including peptide and metabolite abundance, gene expression values, and quantitative microscopy measurements. As a byproduct of this work, sequencing data for these strains was improved from the initial Sanger scaffolds to 30X whole genome coverage and some adjustments to the Liti phylogeny were proposed. With these improved genome sequences and associated phenotypic data, it was possible to draw genotype to both phenotype

and environment connections, e.g. correlating differences in gene expression related to fermentation with strains clustered in fermentation-related ecological niches.

Several of the Liti strains were resequenced to higher coverage not only by Skelly *et al* but by Bergstrom *et al* (Bergström et al. 2014), who used natural isolates to explore questions of the impact of different types of genetic variation on phenotypic variation in different backgrounds, focusing away from single nucleotide variants and onto indels and copy number variation. Previous work with natural isolates had revealed that despite lower genetic divergence, strains of *S. cerevisiae* exhibit more phenotypic variation than strains of more genetically diverged *S. paradoxus*, which was previously thought to relate to the broader occupation of varying ecological niches by *S. cerevisiae* (Liti et al. 2009). Bergstrom *et al* recorded higher incidence of types of genetic variation other than SNVs in *S. cerevisiae*, to which they attribute its greater phenotypic diversity, building on ideas espoused by Legras, Salmon, and others that variation in gene dosage could facilitate rapid response to a changing environment (Legras et al. 2007; Salmon 1997). These findings exemplify the power of studying the genotype-to-phenotype relationship across many strain backgrounds and species to understand the functional outcome of variation at a level beyond what is possible with laboratory strains.

As the value of working with natural isolates has become clear, the number of natural isolates under active study by many different groups has increased. Using a set of 85 *S. cerevisiae* natural isolates (69 wine strains and additional strains used in baking, brewing, and ethanol fuel production), Dunn *et al* identified extensive variation in copy number and transposable element content, as well as introgression events from *S. paradoxus and S. mikitae* and additional genomic regions not even present in the standard reference genome constructed from laboratory strain S288C (Dunn et al. 2012). The pursuit of a wider variety of natural

isolates with associated high quality whole genome sequencing has led ultimately to the 1,002 Yeast Genomes project, now at 1,011 genomes and counting, and to studies that have intentionally enriched for yeasts from underrepresented ecological niches, including coffee and chocolate (Ludlow et al. 2016).

1.2.3 *Applications of natural isolates: mapping studies and complex traits*

Natural isolates provide access to phenotypes and genotypes not observed in lab strains, and large datasets like the "phenomics" dataset provide opportunities to look across strains for variation associated with particular traits and extend the reach of QTL mapping studies. Data from the phenomics study itself facilitated broad inquiries into the prevalence of *cis* eQTLs using association mapping, and demonstrated correlations between *cis* eQTLs and observable changes in phenotype (Skelly et al. 2013). More targeted studies using crosses between two natural isolates, rather than between a natural isolate and a laboratory strain, facilitate the exploration of specific traits with variation in wild yeasts depending on their ecological origin. For example, in an early sporulation efficiency study from Gerke *et al* (Gerke, Chen, and Cohen 2006), a cross between natural isolates revealed that only one of three QTLs identified from crosses between lab strains alone (low efficiency S288C and high efficiency SK1) (Deutschbauer and Davis 2005) was linked to sporulation efficiency in natural isolates. For de-convolving the results of studies like these, knowledge gained from previous work with laboratory strains provided useful guideposts, but further study with natural isolates demonstrated which of the things learned in the lab carry over into the wild.

One subset of phenotypes explored by natural isolates that are not necessarily observed in the lab are those contributing to yeast pathogenicity. Common beer and bread yeast *Saccharomyces cerevisiae* is rarely pathogenic, but in severely immunocompromised patients

can cause infection. Comparative study of and QTL mapping with non-virulent and virulent strains of *S. cerevisiae* has provided a framework for understanding the loci contributing to virulence (Clemons et al. 1994), and to the phenotypes that can assist pathogenic yeasts in the colonization of human substrates (Granek et al. 2013; Steinmetz et al. 2002). Analyzing clinical isolates provides a method for assessing the genetics specifically underlying biofilm formation in the human body, and analyzing biofilm-related phenotypes in other natural isolates will provide insight into the mechanisms they use to colonize other environmental niches.

1.3 BIOFILM-RELATED PHENOTYPES ARE COMPLEX TRAITS WITH MEDICAL AND INDUSTRIAL RELEVANCE

Phenotypes related to yeast biofilm formation are the type of complex trait we can use natural isolates to better understand. Biofilms are communities of cells contained in a protein- and sugar-filled extracellular matrix (Beauvais et al. 2009), adhering to a surface through interactions with cell wall proteins. A subset of biofilm-related phenotypes also include cell-cell adhesion ("flocculation"), instead of or in addition to cell-surface adhesion, also mediated by cell wall protein interactions (Guo et al. 2000). The function of such a community is most often protective: yeast cells exposed individually to a harsh environment, e.g. a chemical treatment, are more likely to die and be unable to propagate their lineage; if they are participating in a community, however, not only are they physically protected by other cells and by the extracellular matrix, it is also more likely that at least some cells from the community will survive to rebuild the population (Vopálenská et al. 2010; Smukalla et al. 2008; Palková 2004; Váchová et al. 2011). The structure of the extracellular matrix also creates pores through which water and nutrients can flow to the cells (Douglas 2003; Váchová et al. 2011). Since many yeast

laboratory techniques rely on the manipulation of single cells or the even distribution of nutrients, biofilm formation is an undesirable trait in the lab. S288C, a widely-used lab strain of yeast, harbors a mutation in a transcription factor involved in biofilm regulation to discourage the development of biofilm-related traits (Liu, Styles, and Fink 1996), and the unstructured growth environments of lab conditions are thought to reduce selection for community growth behaviors (Šťovíček et al. 2010). Although other common lab strains retain their biofilm-related phenotypes, natural isolates are likely to provide an extensive untapped reservoir of these traits.

1.3.1 *Biofilm-related traits contribute to survival and adaptation in natural isolates*

Natural isolates are likely to harbor biofilm-related traits because of the adaptive function they serve in their ecological niches (reviewed in Bauer, Govender, and Bester 2010). Invasive and filamentous growth is thought to be a mechanism for yeast to forage and survive in new environmental niches and sources of nutrients, like rotting fruit (Sarode et al. 2014; Cullen and Sprague 2000; Palecek, Parikh, and Kron 2000), and pathogenic yeasts like *Candida albicans* and *Candida glabrata* use the formation of biofilms and filamentous extensions to colonize human tissues and persist on the surfaces of medical devices such as catheters (Verstrepen and Klis 2006; Cormack, Ghori, and Falkow 1999; Douglas 2003). In *S. cerevisiae*, cell-to-cell adhesion is mediated by interactions between lectin-like cell wall proteins in one cell and the cell wall sugars in another (reviewed in Verstrepen and Klis 2006); in *C. glabrata*, which causes as much as 15% of fungal infections, cell wall proteins in the yeast interact with cell wall sugars in human epithelial cells to allow the yeast to directly colonize human tissues using a biofilm-based mechanism (Cormack, Ghori, and Falkow 1999). The protective environment provided by the biofilm makes these infections difficult to eradicate, and failure to eradicate an infection completely with multiple drug treatments can seed dangerous drug-resistant populations

(reviewed in Ribeiro et al. 2016). *Candida* biofilms resistant to multiple different common antimicrobial treatments are found on implanted catheters and artificial heart valves (Douglas 2003). Drug-resistant fungal infections facilitated by biofilms are a major public health concern of our time. Although *S. cerevisiae* is rarely a human pathogen itself, due to its biofilm-forming ability and genetic homology with more commonly pathogenic organisms like *C. albicans* and *C. glabrata* it is being developed as model for biofilm formation (Reynolds and Fink 2001). Using a model like *S. cerevisiae* to achieve a better understanding of the breadth of biofilm related phenotypes and their underlying genetics could lead to better methods for testing antimicrobial compounds and designing targeted therapeutics (Ribeiro et al. 2016).

Far from its role as an occasional human pathogen, yeast *S. cerevisiae* also uses biofilm-related traits to fulfill one of its most important industrial functions: brewing. Flocculation is desirable in brewing strains, because complete aggregation of the yeast at the end of a fermentation facilitates its easy removal and prevents the development of undesirable flavors (reviwed in Verstrepen et al. 2003). It has even made its way into common vernacular, with "top fermenters" describing ale yeasts that rise upon flocculating, and "bottom fermenters" describing lager yeasts that sink (Verstrepen et al. 2003). The same traits that make yeast useful for brewers are also useful in other areas of industry. Yeast are used to ferment ethanol for fuel (bioethanol) and at the end of a fermentation are extracted expensively from the culture using centrifugation or chemical immobilization, processes that natural sedimentation due to flocculation would render unnecessary; flocculent yeast are therefore being considered a new technology to increase the efficiency and decrease the expense of bioethanol production (reviewed in Sivakumar et al. 2010). An improved understanding of yeast flocculation and the spectrum of phenotypes

available in natural isolates is important for selecting and designing yeast to use in industrial bioreactors (reviewed in Domingues et al., 2000).

1.3.2 *The physical and genetic basis of specific biofilm-related traits*

Yeast express a variety of cell wall proteins for the purpose of forming different types of biofilms, with different primary cell wall protein contributors. We know now from numerous genetic studies that these lectin-like cell wall proteins or "flocculins" are produced by genes in the *FLO* family, composed of *FLO1, FLO5, FLO10,* and *FLO11* (Verstrepen and Klis 2006; Guo et al. 2000). Flocculin proteins mediate both the cell-cell and cell-surface adhesion properties that are characteristic of biofilm-related traits and contribute to phenotypes including the formation of colonies and mats with complex morphology, flocculation, invasion into agar, and adhesion to plastic surfaces.

Complex colony morphology: Yeast growing on solid agar are organized into colonies, which can have a variable morphology depending on strain background, nutrient availability, cell shape, budding pattern, and cell surface proteins (Vopálenská et al. 2005). Strains with the capability to generate an extracellular matrix frequently form more intricate colonies than the smooth, featureless colony generated by laboratory strain S288C (Vopálenská et al. 2005); these filigreed colonies are often described as "fluffy" or complex, and are also called "biofilm colonies" due to the biofilm properties they exhibit (Váchová et al. 2011). Most of what is known about complex colonies comes from the study of natural isolates of yeast, which frequently form complex colony morphologies that are lost through domestication (Šťovíček et al. 2010), although laboratory strains have been observed to form complex colonies under nutrient limited conditions (Gimeno et al. 1992; Vopálenská et al. 2005). Under repeated growth

conditions, colony morphology is replicable within a given strain background, indicative of a genetically-specified development program (Vopálenská et al. 2005). This trait is grouped as a biofilm-trait due to the presence of an extracellular matrix in complex colonies, verified by Stovicek *et al* in 2010 (Šťovíček et al. 2010).

Characterization of smooth and complex colonies has revealed that despite a greater wet biomass of complex colonies, smooth colonies contain a greater number of cells overall, a difference that has been shown to be due to the presence of high extracellular matrix content in complex colonies that retains extracellular fluid. The amount and contents of this matrix, when extracted by centrifugation and analyzed for characteristic glycoprotein concentrations, is variable between strains with different morphologies. Stovicek *et al* validated these findings with gene expression comparisons, identifying flocculin gene *FLO11* and water channel gene *AQY1* as highly expressed in complex colonies. Further experiments deleting *FLO11* caused a complete collapse of the 3D colony structure (Šťovíček et al. 2010). Variation in gene expression in cells throughout the colony suggests that a thick layer of surface cells express genes encoding multi-drug resistance plasma membrane transporters that facilitate the removal of toxins and protect the colony interior, and certain small chemical species are prevented from reaching cells at the colony base (Váchová et al. 2011).

Complex morphology in *C. albicans* is characterized by phenotypic switching between simple and complex colony forms to respond to environmental signals, and a distribution of cells with different morphologies (round, filamentous) throughout. This "dimorphic transition" between round, apically budding and filamentous, unipolar budding cells also occurs in the formation of complex colonies by laboratory strain Σ1278B, starting on the edges of developing microcolonies (Vopálenská et al. 2010). This switch occurs in response to signaling between

microcolonies via the production of ammonia, and can be induced by volatile ammonia and associated alkalization of the media. The colonies grow smoothly until a second dimorphic switch causes wrinkles to form on the surface, the result of a subset of cells converting to filamentous, pseudohyphal growth, which develop in a regular pattern (Vopálenská et al. 2010). Flo11 localization studies revealed that Flo11 is not needed for the first dimorphic switch, but for the second dimorphic switch to establish the complex final structure, and that Flo11 is also needed for attaching the colony to the agar surface, classifying it as a biofilm (Vopálenská et al. 2010). More detailed analyses of the internal structure of complex colonies suggest that Flo11 protein molecules actually function as anchoring fibers participating in the physical structure of the colony (Váchová et al. 2011).

The majority of colony morphology studies have treated it as a qualitative, rather than quantitative, complex trait described based on phenotype, or scored for complexity from 1-5 as a proxy for quantitation (Granek et al. 2013). Recent work, however, digitally quantified aspects of colony complexity using an automated image capture platform and extracting image features to describe the general features including shape, texture, and intensity. Additional automated feature extraction yielded a final list of 427 separate features which were used to build profiles of colonies as they grow over time and ultimately in a model to automatically identify fluffy or smooth colonies (Ruusuvuori et al. 2014). Not only do automated methods like this provide a platform for better characterizing the variation we observe in natural isolate phenotypes, but they provide a framework for the type of high-throughput phenotyping needed for any mapping or variant library study.

Complex mat formation: A different complex biofilm phenotype can be observed when a colony is inoculated on a plate with lower agar concentration, and this phenotype is thought to be

related to a bacterial behavior. When bacterial cells are grown on an agar plate with a low concentration of agar, they engage in a behavior called “sliding motility”, resulting from hydrophobicity of cell surface proteins causing reduced friction between the expanding bacterial population and their growth substrate (Reynolds and Fink 2001). Reynolds and Fink examined yeast for a similar behavior and found that yeast grown on low agar plates form large, spreading mats rather than structured colonies (Reynolds and Fink 2001). Mats can exhibit a complex filigreed or spoke-and-petal morphology specific to the strain background (Reynolds et al. 2008), with fine surface features formed by water channels (Sarode et al. 2014) and a smooth leading edge (Reynolds et al. 2008). The edge, or rim, of the mat exhibits different properties from the internal hub, including active growth and reduced adherence to the substrate thought to be caused by a pH gradient (Reynolds et al. 2008). Microscopy of cells in mats formed by biofilm-competent lab strain Σ1278B revealed that all of the mat cells appear to be in round “yeast form” rather than filamentous, despite the dual role of *FLO11* in filamentous growth and in mat formation, for which it is required (Reynolds and Fink 2001). Reynolds and Fink also verified that cells expressing *FLO11* have increased hydrophobicity compared to null mutants, supporting the sliding motility hypothesis underlying this spreading behavior.

In addition to a functional Flo11 protein, complex mat formation also depends on functional signaling pathways, as strains with null mutations in signaling pathway components all exhibit reduced speed and complexity of mat formation, thought to be due to the importance of glucose sensing in the development of this phenotype (Reynolds and Fink 2001; Reynolds et al. 2008). Whole genome transcriptional comparison between planktonic and mat-forming cells has revealed an additional important interaction between transcription factor Opi1 and Flo11 (Reynolds 2006). Additional work has identified interacting proteins Wsc1 and Rom2 and their

target Skn7 as essential factors for correct mat formation as well. These proteins are part of a cell wall integrity pathway and participate in mat formation independently from Flo11; they are also not essential for any other Flo11 mediated biofilm phenotypes (e.g. invasion) (Sarode et al. 2014).

Agar invasion: Initially described by Roberts and Fink in 1994, haploid invasive growth is related to the filamentous growth program of diploid strains and characterized by the formation of long pseudohyphal cells that penetrate agar media substrates. Round, axially budding cells switch to a polar budding pattern to facilitate filament formation in response to MAPK signaling, a switch which enables them to grow into the agar underneath a colony (Roberts and Fink 1994). This invasion is observed after colonies are washed from the surface of the plate, and can be validated as true invasion if the agar must be penetrated to access the cells (Roberts and Fink 1994). The switch in budding pattern can be induced by environmental signals, particularly low glucose availability, supporting the hypothesis that invasive growth is a nutrient foraging behavior (Cullen and Sprague 2000; Palecek, Parikh, and Kron 2000). There is evidence that cells participating in invasive growth express enzymes to degrade pectin and starch in the actual substrate (Palecek, Parikh, and Kron 2000).

The role of transcription factor Flo8 in invasive growth was described by Liu *et al*, who identified the *FLO8* stop codon in lab strain S288C that prevents it from manifesting most biofilm-related phenotypes (Liu, Styles, and Fink 1996). In crosses, *FLO8* segregates perfectly with invasive ability, and transformation of S288C with a working *FLO8* from Σ1278B largely but not completely restores haploid invasive growth (which other genetic factors could attenuate in this background) (Liu, Styles, and Fink 1996). Invasive ability is also lost in null mutants of *flo10* and *flo11* but not *flo1*, and overexpression strains demonstrate increased invasive growth

with higher expression of *FLO10* and *FLO11* (Guo et al. 2000); one of the earliest known functions of *FLO11* was that it was required for invasive growth in biofilm-competent strains (Lo and Dranginis 1998). Key members of the MAP kinase signaling pathway are also required for invasive growth in lab strain S288C (Liu, Styles, and Fink 1996; Lo and Dranginis 1998), connecting it to the environmental responsiveness that characterizes this and other biofilm-related traits. Due to the length of the *FLO11* promoter (the longest in the yeast genome) and the myriad factors functioning to activate or repress its activity, as well as the potentially separate control of cell adhesion and budding polarity (Palecek, Parikh, and Kron 2000; Mösch and Fink 1997), the complete picture of genes involved in invasive growth remains to be untangled (Palecek, Parikh, and Kron 2000).

In the literature, there is frequently an unclear separation between invasion into agar and adhesion to agar (Guo et al. 2000), and surface adhesion is best separately characterized by examining adhesion properties on another substrate.

Surface adhesion: The study of static adhesion to a plastic substrate was initially developed for bacterial biofilms and has been adapted to use with yeast (O'Toole, Kaplan, and Kolter 2000). Bacteria or yeast adhering to an abiotic substrate, usually plastic, are stained with crystal violet, a dye that has traditionally been used to stain bacteria and identify them as gram-positive (Beveridge 2001) but can stain any cell wall (Popescu and Doyle 1996). In biofilm assays, staining facilitates visualization of the biofilm attachment pattern and, if the dye is subsequently solubilized, quantification of the amount of stain absorbed (Merritt, Kadouri, and O'Toole 2005). Analyses of cells forming biofilms on abiotic surfaces have revealed the production of extracellular polysaccharides (O'Toole, Kaplan, and Kolter 2000) and other polymeric

components of an extracellular matrix (Baillie and Douglas 2000) making the descriptor "biofilm" an accurate characterization of this community growth behavior.

For high-throughput screening of potentially biofilm-capable strains, a polystyrene 96-well plate is typically used as the substrate (Merritt, Kadouri, and O'Toole 2005; O'Toole, Kaplan, and Kolter 2000). Initially developed for use with *C. albicans* (Ramage et al. 2001), it has been adapted for use with *S. cerevisiae* as a new model for fungal biofilm formation (Reynolds and Fink 2001). Deletion of *FLO8* (the regulator of flocculin gene *FLO11*), and of *FLO11* itself each was sufficient to eliminate the ability of biofilm-forming laboratory strain Σ1278B to adhere to polystyrene (Reynolds and Fink 2001). Reynolds and Fink also discovered that glucose and therefore active metabolism is also required for adhesion, although the adhesive properties are increased at low glucose or limited nutrient concentrations.

This 96-well screening technique has been useful for screening the biofilm-forming capabilities of many strains, including an effort from Li *et al* to characterize biofilms formed by 115 natural isolates of *C. albicans*, 84 of which were clinically derived (Li, Yan, and Xu 2003). Analysis of these isolates with microscopy and crystal violet stain quantification revealed extensive quantitative variation in biofilm-forming ability, with a more than 50-fold difference between the weakest and strongest biofilm-forming strains; selective genotyping also revealed genetic diversity, even among samples from the same clinical origin. Biofilm-forming ability was correlated with measures of relative hydrophobicity, a trait thought to facilitate cell-surface interactions in biofilms (Reynolds and Fink 2001). Working with natural isolates, even in a simulated environment, can reveal a broad spectrum of phenotypes not observed in lab strains. Biofilm formation can also be studied in assays designed to mimic real scenarios.

For investigation of biofilm-forming properties in simulated real systems, yeast can be grown on the same material from which catheters are made, or screened for their ability to grow on a variety of common substrates used in medical settings (O'Toole, Kaplan, and Kolter 2000; Baillie and Douglas 1999). Baillie and Douglas used microscopy to characterize the attachment patterns of *C. albicans* biofilms on catheter plastic and revealed a role for switching between round yeast and filamentous forms, with round cells most effectively forming the initial attachment and hyphal cells forming an additional layer on top. This phenotype changed in mutant backgrounds and on different substrates, but all of the biofilms in this study were resistant to antifungal agent amphotericin B (Baillie and Douglas 1999).

Flocculation: Though the majority of classic biofilm-related traits concern adhesion between cells and surfaces, cell-cell adhesion, or flocculation, is an alternative type of biofilm that affords participating cells strong protection (Smukalla et al. 2008; Beauvais et al. 2009) and can also occur in response to limited nutrients, especially sugars (Guo et al. 2000). Flocs also do incorporate an extracellular matrix. The components of an extracellular matrix can be different in each organism and strain, and also across phenotypes. Beauvais *et al* identified a polysaccharide-rich extracellular matrix in flocs from a flocculent wild isolate of *S. cerevisiae*, validating the inclusion of flocs as a biofilm-related phenotype, and demonstrated that its production depends on a functional *FLO1* (Beauvais et al. 2009). Electron microscopy revealed a visible electron-dense layer on the cellular surface of flocculent cells that could be removed with multiple cycles of deflocculation treatment, and analysis of the extracted material confirmed it to be the hexose-rich extracellular matrix produced by the flocculating cells. Interestingly, *FLO1*-expressing cells without an extracellular matrix do still floc, indicating its primary function might be protective. Indeed, the barrier it creates around the cell walls of individual cells participating in flocs

prevents large molecules including antibodies from accessing the cell, but does not contribute to the resistance flocs demonstrate to chemical treatments (Beauvais et al. 2009). Protection from chemical treatments has been convincingly demonstrated to arise from the tightly-packed structure of the floc, which protects cells on the interior; these interior cells undergo chemical changes promoting stress resistance (Smukalla et al. 2008).

The primary flocculin protein contributing to flocculation is Flo1, essential for the formation of flocs and of an extracellular matrix (Smukalla et al. 2008; Beauvais et al. 2009). Not only is its expression tightly correlated with flocculation in natural isolates, inducible expression of *FLO1* also causes non-flocculent lab strain S288C to flocculate (Smukalla et al. 2008), and overexpression enhances the phenotype in already flocculent strains (Guo et al. 2000). Introduction of a wild-type *FLO8* into *flo8* non-flocculent strains is sufficient to induce flocculation (Liu, Styles, and Fink 1996), and a mapping study linking *FLO1* to flocculation as a quantitative trait also validated the role of *FLO8* (Brem et al. 2002). Additional work has demonstrated that Flo5 proteins (along with Flo1 proteins) are also present at the cell surface in flocs and correlated with flocculation amount (Bony, Barre, and Blondin 1998); other studies also suggest that in certain strain backgrounds and conditions, Flo11 can contribute to flocculation as well (Bayly et al. 2005). This work is further validated by overexpression studies in a flocculent strain, which demonstrate the involvement of *FLO10* and *FLO11* in flocculation (in the background examined) (Guo et al. 2000). Other non-*FLO* genes show changes in expression in cells participating in flocs, including *DAN1*, which encodes a cell wall protein involved in anaerobic growth, which might also play a role in resistance to certain drugs (Smukalla et al. 2008).

1.3.3 *Biofilms as complex and quantitative traits*

Common features of complex traits include contributions by multiple genes to an individual phenotype and influence of the phenotype by environmental factors. By these criteria, each of these five primary biofilm-related traits is complex, and with the exception of mat formation, quantitative.

Although the role of the *FLO* genes and flocculin proteins in biofilm-related traits is well characterized, several deletion collection and mapping studies focused on the traits have identified many additional genes contributing to the complexity of these traits. Filamentous growth, a trait related to haploid invasion, is among the most studied of biofilm-related traits via screening with deletion and mutagenesis collections. One of the earliest studies of this nature, a transposon mutagenesis screen in 1997, identified 16 new mutants defective for filamentous growth, which were distributed in overlapping categories depending on their involvement in cell polarity, cell shape, and/or invasive growth; the novel genes identified included *STE12* and *TEC1*, now known to be important players for determining whether invasive growth occurs (Palecek, Parikh, and Kron 2000). Separate studies in 2008 and 2012 used transposon insertion, overexpression, and deletion collections made in Σ1278B (Jin et al. 2008; Ryan et al. 2012). Analysis of the transposon insertion and overexpression collections yielded 487 total genes (including the *FLO* genes) with an effect on filamentous growth, with roles including cell morphogenesis, cell growth, filamentous growth, and nitrogen utilization, as described by GO term analysis (Jin et al. 2008). The study of filamentous growth using the Σ1278B deletion collection revealed 700 contributing genes (Ryan et al. 2012). The Σ1278B deletion collection has also been used to exhaustively identify genes involved in haploid invasive growth (577) and mat formation (688) (Ryan et al. 2012), as well as in complex colony formation (211)

(Voordeckers et al. 2012) which include genes spanning all of the major cellular signaling pathways. The vast number of genes identified in screens of this nature is indicative of the complexity of such environmentally responsive traits, but also fails to reveal the relative effect size or interactions among identified genes.

QTL mapping has also presented a successful strategy to identify strong effect loci contributing to these traits, including but not limited to the *FLO* family of genes. In Bulk Segregant Analysis of flocculation in a cross between non-flocculent lab strain S288C and flocculent lab strain SK1, *FLO8* was identified within a major effect QTL (Brauer et al. 2006). Multiple mapping analyses of complex colony morphology using clinical yeast isolates have revealed QTLs containing *FLO8, FLO11* and known components of cell signaling pathways, but also causal genes involved in oxidative stress response and endocytosis, as well as important contributing cryptic variants (Granek et al. 2013; Granek and Magwene 2010; Taylor and Ehrenreich 2014; Taylor et al. 2016) which reinforce the background dependence of these traits.

Many of the genes identified in these deletion, mutagenesis, and mapping studies are involved in signaling pathways that notify yeast of changes in its environment. The ability of yeast to modify its cell surface properties to respond to environmental conditions is an important adaptive tool, and biofilm-related traits are frequently observed as that mechanism for adaptation to environmental stimuli. The genes that encode the essential flocculin proteins that facilitate biofilm-related traits have lengthy promoters regulated by the inputs of most major cellular signaling pathways, and the role of nutrient limitation in enhancing major biofilm-related phenotypes is well documented. This clear relationship between biofilm-related traits and environmental stimuli makes it even more important to study these traits and their underlying

genetics in strains that have been subjected to a variety of environmental pressures in different ecological niches.

1.3.4 *Natural isolates provide access to new phenotypes and genotypes*

The workhorse of yeast biofilm research in *S. cerevisiae* is flocculent, biofilm-forming laboratory strain Σ1278B: most of what we know and guess about the genetics and complex characteristics of these traits comes from the study of this single, well-characterized strain. Early forays into work with natural isolates have revealed novel phenotypes, as with complex colony morphology studies, and there is evidence that certain genetic truths, like the role of *FLO11* in cell-surface but not cell-cell adhesion, are highly background-dependent (Bayly et al. 2005; Guo et al. 2000). By studying traits that specifically have to do with environmental response and exploration in laboratory strains, the scope of our work has been limited. By working with natural isolates, with phenotypes shaped by their particular ecological niches, we can begin to understand the full spectrum of biofilm-related phenotypes available. This added dimension from natural isolates, combined with the insights provided by experimental evolution of laboratory strains, is a powerful tool to understand the full complement of genetic contributions to these traits.

1.4 SPECIFIC AIMS

1.4.1 *Leveraging natural isolates to understand the spectrum of complexity of biofilm-related traits*

It is clear from the existing body of work with natural isolates that they are powerful tools for understanding the breadth of complex phenotypes for industrially relevant traits, and for testing hypotheses generated with laboratory strains for validity across divergent backgrounds. Natural

isolates are expected to exhibit extensive diversity in the phenotypes related to biofilm formation, including cell-cell and cell-surface adhesion traits. However, despite their medical and industrial importance, they have yet to be exhaustively characterized. Furthermore, the contributions of the types of genetic variation frequently seen in natural isolates, including ploidy variation, are unknown outside a laboratory setting but thought to be important (Zörgö et al. 2013). The first of my aims is designed to quantify and describe the biofilm-related phenotypes achieved by the Liti natural isolates (Liti et al. 2009) and determine how they are affected by changes in ploidy.

1.4.2 *Leveraging experimental evolution to identify favored and novel genetic contributors to flocculation*

Many of the traits that determine the pathogenicity or industrial utility of a microorganism like yeast are controlled by more than one gene. However, not all relevant genes may be equally important, and existing methods for determining the genetic basis of these complex traits identify many contributing genes with no sense of whether they represent viable adaptive routes in a competitive environment. Experimental evolution has been used to reveal the most successful evolutionary routes underlying many complex traits, including nutrient limitation and ethanol tolerance (Brown, Todd, and Rosenzweig 1998; Hong and Gresham 2014; Voordeckers and Verstrepen 2015). We examine how a single biofilm-related phenotype, flocculation, evolves in continuous culture, with the final goal of engineering a strain of yeast with low adaptive potential for flocculation in the laboratory. The cycle of evolution, sequencing, and design we introduce in this study provides a framework for using experimental evolution to reduce the occurrence of other undesirable traits.

There are myriad future directions on the interface between these projects, combining the adaptive potential of different background strains with the diversity supplied by natural isolate collections.

Chapter 2. PLOIDY-REGULATED VARIATION IN BIOFILM-RELATED PHENOTYPES IN NATURAL ISOLATES OF *SACCHAROMYCES CEREVISIAE*

The work described here was performed independently by Elyse Hope and is derived from:

Hope, Elyse A., and Maitreya J. Dunham. 2014. "Ploidy-Regulated Variation in Biofilm-Related Phenotypes in Natural Isolates of *Saccharomyces cerevisiae*." *G3: Genes|Genomes|Genetics* 4 (9): 1773–86. doi:10.1534/g3.114.013250. doi: 10.1534/g3.114.013250

The ability of yeast to form biofilms contributes to better survival under stressful conditions. We see the impact of yeast biofilms and "flocs" (clumps) in human health and industry, where forming clumps enables yeast to act as a natural filter in brewing and forming biofilms enables yeast to remain virulent in cases of fungal infection. Despite the importance of biofilms in yeast natural isolates, the majority of our knowledge about yeast biofilm genetics comes from work with a few tractable laboratory strains. A new collection of sequenced natural isolates from the *Saccharomyces* Genome Resequencing Project enabled us to examine the breadth of biofilm-related phenotypes in geographically, ecologically, and genetically diverse strains of *Saccharomyces cerevisiae*. We present a panel of 31 haploid and 24 diploid strains for which we have characterized six biofilm-related phenotypes: complex colony morphology, complex mat formation, flocculation, agar invasion, polystyrene adhesion, and psuedohyphal growth. Our results show there is extensive

phenotypic variation between and within strains, and that these six phenotypes are primarily uncorrelated or weakly correlated, with the notable exception of complex colony and complex mat formation. We also show that the phenotypic strength of these strains varies significantly depending on ploidy, and the diploid strains demonstrate both decreased and increased phenotypic strength with respect to their haploid counterparts. This is a more complex view of the impact of ploidy on biofilm-related phenotypes than previous work with laboratory strains has suggested, demonstrating the importance and enormous potential of working with natural isolates of yeast.

2.1 INTRODUCTION

Wild type budding yeast selectively bind to each other and to substrates in cohesive groups called yeast "flocs" (or clumps) and yeast biofilms. The ability of yeast to form biofilms contributes to better survival under stress and affects myriad processes important to human health and industry. In brewing, yeast flocs facilitate easy removal of yeast from the final product, but yeast biofilms can complicate biofuel production and help pathogenic yeast remain virulent on hospital surfaces and during infection. In the laboratory, yeast biofilms create significant challenges for many experiments. As a result, biofilm formation has been selected against in most laboratory strains and our knowledge of the genetic basis of yeast biofilm formation remains incomplete.

Yeast flocs and biofilms are two distinct but related phenotypes regulated by environmental changes that cause complex signaling and gene expression responses. The formation of a true biofilm requires yeast to form an extracellular protein matrix and adhere to a surface. The matrix is made up of secreted proteins and the biofilm is anchored to surfaces by hydrophobic interactions between yeast cell surface proteins and the surface (Voordeckers et al. 2012). Biofilms can be observed phenotypically by several assays including examination of a

strain's ability to adhere to polystyrene surfaces, to form a complex colony with an extracellular matrix, or to form complex mats on low agar media (O'Toole and Kolter 1998; O'Toole, Kaplan, and Kolter 2000; Reynolds and Fink 2001; Granek and Magwene 2010; Šťovíček et al. 2010). Flocs are clumps of yeast that adhere to each other by a protein-carbohydrate bond between cell surface proteins and cell wall sugars (Bony, Barre, and Blondin 1998; Guo et al. 2000; Verstrepen et al. 2003; Brückner and Mösch 2012). This phenotype is most commonly observed as clumping and settling in liquid media (Liu, Styles, and Fink 1996; Smukalla et al. 2008; Johnston and Reader 1983). Due to the intercellular bonds, flocculation is different from the type of clumping that results from mother-daughter cell separation defects. Flocculent clumps have been shown to render the participating yeast cells more resistant to chemical stresses (Smukalla et al. 2008). There is also evidence that flocs contain an extracellular matrix, a biofilm-related trait, reflecting the close relationship between biofilm and floc phenotypes (Beauvais et al. 2009). The formation of biofilms and flocs are thought to be regulated by similar genes. The most well-studied of these genes are the *FLO* genes, including the important genes *FLO1* and *FLO11*, involved primarily in cell-cell and cell-surface adhesion, respectively, and transcription factor *FLO8*, which regulates expression of the other *FLO* genes (Verstrepen and Klis 2006).

Biofilm- and flocculation-related phenotypes are predicted to differ depending on ploidy. Another common phenotype associated with *FLO* gene activity is invasiveness into solid agar media, which in haploids occurs on rich media and in diploids occurs under limited nitrogen supply as filamentous growth (Liu, Styles, and Fink 1996; Cullen and Sprague 2000; Guo et al. 2000; Ryan et al. 2012). Diploid strains under a limited nitrogen supply are able to undergo a dimorphic switch from a yeast form, which generates colonies on budding, and the filamentous form, which generates long chains of interconnected cells called "pseudohyphae" that can be

seen as filamentous extensions from a colony (Gimeno et al. 1992; Liu, Styles, and Fink 1996; Lo and Dranginis 1998). Increases in ploidy have also previously been shown to reduce complex colony morphology, complex mat formation, and invasion; this decreased phenotypic strength is thought to result from a cellular response to increased gene dosage or DNA content, or differential expression of cell-surface proteins (Galitski et al. 1999; Guo et al. 2000; Reynolds and Fink 2001; Granek and Magwene 2010). The possible phenotypic differences between haploid and diploid versions of the same strain, as well as the filamentous growth phenotype specific to diploid strains, make a compelling case for considering diploid as well as haploid versions of a strain in any effort to generate a complete picture of its biofilm-forming capabilities.

Natural isolates provide a unique testing ground for understanding the spectrum of biofilm-related phenotypes achieved by *S. cerevisiae* outside of laboratory strains.
The existence of a recent collection of natural yeast isolates from the *Saccharomyces* Genome Resequencing Project (Liti et al. 2009) makes it possible to examine important phenotypes that exist among natural yeasts but are not prevalent in laboratory strains. It also provides an opportunity to explore the connection between genotype and biofilm-related phenotypes on an unprecedented scale. The SGRP collection of natural isolates continues to grow, providing an extended resource for examining questions of phenotypic variation in natural isolates and now, with high-quality sequence information available, questions of genotype to phenotype connections (Skelly et al. 2013; Bergström et al. 2014).

Other recent studies have used the SGRP collection and other similar natural isolate collections (e.g., Phaff Yeast Culture Collection, ARS (NRRL) Culture Collaboration) to examine questions of lineage and sequence similarity across geographical niches (Liti et al.

2009; Cromie et al. 2013). The SGRP collection also has associated "phenome" data, for which each strain's morphology, metabolite, gene expression, and protein traits were recorded, as well as the strains' responses to environmental perturbations including toxins and nutrient limitations (Zörgö et al. 2013; Warringer et al. 2011; Skelly et al. 2013). With this extensive profile available on the sequence, gene expression, and phenotypic levels, the SGRP collection is a natural and informative starting point for examining genotype/phenotype relationships, and strains from the collection have already been used effectively in a few quantitative trait mapping studies (Cubillos et al. 2011; Cubillos et al. 2013; Granek et al. 2013; Jara et al. 2014).

Our objective in this study was to survey the diversity of biofilm-related phenotypes in the natural isolate collection and to determine the extent to which any or all of those biofilm-related phenotypes are related. We present a panel of biofilm-related phenotypic assays for 31 haploid and 24 diploid strains in the SGRP natural isolate collection. We show that there is strong qualitative and quantitative variation across six primary phenotypes within the collection: complex colony morphology, complex mat formation, flocculation, agar invasion, polystyrene adhesion, and filamentous growth. Our results show that, with some notable exceptions, the strength of each phenotype is only weakly correlated with the others, demonstrating the potential of these assays to evaluate different aspects of biofilm-related traits. We also show that the diploid phenotypes are significantly different from their haploid counterparts, in several cases showing increased complexity with respect to the haploid strains contrary to expectations based on the literature about laboratory strains. These findings clearly demonstrate the utility of examining natural isolates to better understand the correlations between these phenotypes and their relationship to ploidy.

2.2 RESULTS

2.2.1 *Natural isolates exhibit extensive diversity across biofilm-related phenotypes*

To determine the diversity of phenotypes present in the *Saccharomyces* Genome Resequencing Project collection (Liti et al. 2009), we examined five different biofilm-related phenotypes across 30 *MATa* haploid strains from the collection and a laboratory strain control (Table 2.1).

Table 2.1 *S. cerevisiae* strains included in this study

NAME	GEOGRAPHICAL ORIGIN	ECOLOGICAL ORIGIN	MATING TYPE	DIPLOID INCLUDED
DBVPG6765	Unknown	Unknown	a	yes
SK1	USA	Soil	a	yes
DBVPG6044	West Africa	Bili wine, from *Osbeckia grandiflora*	a	yes
DBVPG1373	Netherland	Soil	a	yes
DBVPG1853	Ethiopia	White Teff	a	yes
Y55	France	Grape	a	yes
YPS128	Pennsylvania, USA	Soil beneath *Q. alba*	a	yes
DBVPG1106	Australia	Grapes	a	yes
DBVPG6040	Netherland	Fermenting fruit juice	a	yes
BC187	Napa Valley, USA	Barrel fermentation	a	yes
YPS606	Pennsylvania, USA	Bark of *Q. rubra*	a	yes
L-1374	Chile	Fermentation from must Pais	a	yes
L-1528	Chile	Fermentation from must Cabernet	a	yes
NCYC361	Ireland	Beer spoilage strain from wort	a	yes
K11	Japan	Shochu sake strain	a	yes
Y12	Africa - Ivory Coast	Palm wine strain	a	yes
YS2	Australia	Baker strain	a	no
YS4	Netherlands	Baker strain	a	no
YS9	Singapore	Baker strain	a	no
UWOPS83-787.3	Bahamas	Fruit, *Opuntia stricta*	a	yes

UWOPS87-2421	Hawaii	Cladode, *Opuntia megacantha*	a	yes
UWOPS05-217.3	Malaysia	Nectar, Bertram palm	a	yes
UWOPS05-227.2	Malaysia	*Trigona* spp (Stingless bee) collected near Bertram palm flower	a	yes
W303	Unknown	N/A	a	no
322134S	RVI, Newcastle, UK	Clinical isolate (throat sputum)	a	no
378604X	RVI, Newcastle, UK	Clinical isolate (sputum)	a	no
273614N	RVI, Newcastle, UK	Clinical isolate (fecal)	a	yes
YJM978	Bergamo, Italy	Clinical isolate (vaginitis)	a	yes
YJM981	Bergamo, Italy	Clinical isolate (vaginitis)	a	yes
YJM975	Bergamo, Italy	Clinical isolate (vaginitis)	a	yes
FY4	Laboratory strain	Laboratory strain	a	no

The five phenotypes (complex colony formation, complex mat formation, flocculation/settling, agar invasion, polystyrene adhesion) are shown for a representative set of strains in Figure 2.1. The full dataset for two biological replicates is provided in Appendix A.1.

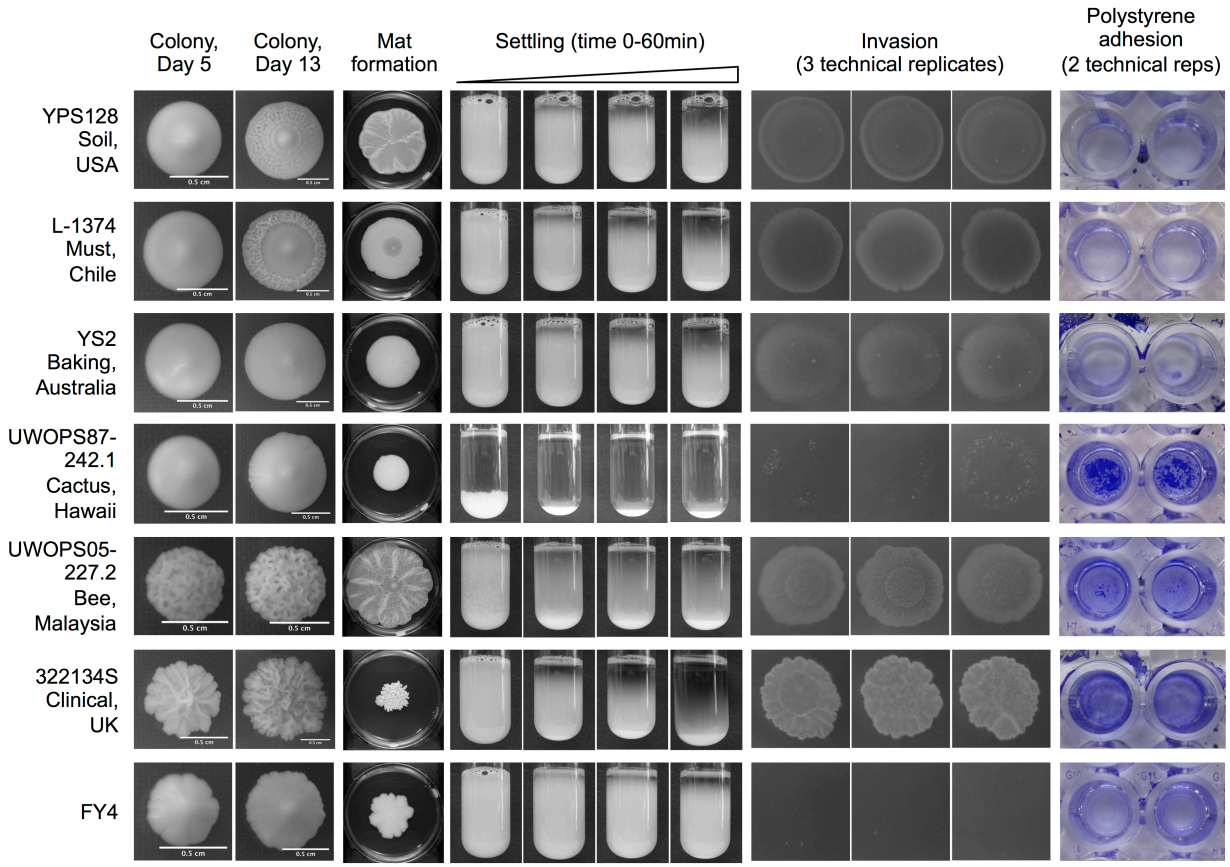


Figure 2.1. Natural isolates exhibit extensive diversity across biofilm-related phenotypes. Full phenotypic panel for six representative strains and reference strain FY4. Strains are listed with their formal name and origin and are shown across five different phenotypes. Complex colony morphology is shown at day 5 and day 13 of growth on 2% agar YPD plates. Complex mat formation is shown at day 13 of growth on 0.3% agar YPD plates with the plate included for scale. Settling photo time points were taken at 0, 15, 30, and 60 minutes. Three technical replicates are shown for the invasion assay, photographed after 24 hours' growth following washing on day 5. Two technical replicates are shown for the polystyrene adhesion assay. Pictured biofilms are fixed and stained with a 1% w/v crystal violet solution.

These six representative strains not only originate from varied ecological and geographical niches, but also demonstrate the full range of phenotypes we observed in this collection from grade 0 (no phenotype) to grade 5 (very strong phenotype) as outlined in Table 2.2, with score assignments shown in Table 2.3.

Table 2.2 Qualitative scoring metrics

ASSAY	SCORE	CRITERIA
Complex colony morphology	0	simple colonies
	1	non-smooth colony surface
	2	signs of complex morphology
	3	moderate complex morphology (does not cover entire colony)
	4	strong complex morphology (covers entire colony)
	5	very strong complex morphology
Complex mat formation	0	simple mat
	1	very light sectoring and/or light ruffle on edge
	2	clear ruffle on edge or petals, can include sectoring
	3	obvious complexity at edges
	4	obvious complexity at edges and in center
	5	strong complexity across entirety of mat
Diploid filamentous growth	0	no filaments
	1	short filaments sparsely distributed around perimeter
	2	long filaments sparsely distributed around perimeter
	3	long filaments distributed around majority of perimeter
	4	moderate length filaments densely distributed around entire perimeter
	5	long filaments densely distributed around entire perimeter

Table 2.3 Qualitative score assignments

STRAIN	HAPLOID BIOREP1 COLONY	HAPLOID BIOREP2 COLONY	HAPLOID BIOREP1 MAT	DIPLOID COLONY	DIPLOID MAT
DBVPG6765	0	0	0	0	1
SK1	4	4	4	4	4
DBVPG6044	2	2	2	1	4
DBVPG1373	0	0	0	0	3
DBVPG1853	1	2	0	1	3
Y55	1	0	2	2	0
YPS128	3	3	4	2	0
DBVPG1106	0	0	0	0	1
DBVPG6040	4	3	4	1	2
BC187	2	2	5	0	1
YPS606	3	3	5	0	0
L-1374	3	3	4	2	2
L-1528	4	5	5	3	3
NCYC361	0	0	0	0	3
K11	3	3	5	3	4
Y12	0	0	0	0	1
YS2	0	0	1		
YS4	1	1	3		
YS9	0	0	2		
UWOPS83-787.3	1	1	0	0	0
UWOPS87-2421	1	1	0	1	0
UWOPS05-217.3	5	5	5	2	4
UWOPS05-227.2	5	5	5	1	4
W303	0	0	2		
322134S	5		5		
378604X	1	1	2		
273614N	1	1	0	0	1
YJM978	3	2	5	0	3
YJM981	5	5	5	3	2
YJM975	0	0	1	2	2
FY4	2	2	3		

Strains are assigned qualitative scores based on the metric described in Table 2.2.

We observed that many strains differed in the strength of each trait. For example, the Hawaiian cactus strain UWOPS87-242.1 exhibits strong flocculation and polystyrene adhesion phenotypes, but weak phenotypes across the other assays (Figure 2.1). Within a single phenotypic grading

there was also significant morphological variation making each strain unique. In the complex mat formation assay for example, strains YPS128, UWOPS05-227.2, and 322134S formed highly complex mats that received qualitative scores of 4 and 5, but all three mats exhibit distinct morphologies (Figure 2.1). By examining the strains using a range of phenotypic assays we were able to generate what is essentially a unique profile of biofilm-related phenotypes for each of the 31 strains, showcasing the incredible diversity accessible in this collection of natural isolates.

2.2.2 *Quantitative measures of biofilm-related phenotypes agree with qualitative scoring*

In addition to the qualitative scores assigned to each phenotype, the flocculation, invasion, and polystyrene adhesion assays were also amenable to quantitation as has been described in previous studies (see Materials and Methods). Previous literature has attempted to assign quantitative metrics to mat formation as well (Reynolds and Fink 2001). Though major mat features were retained across biological replicates in our dataset (e.g., Figure 2.2 D), quantifiable features like spoke number and lobe number were not consistent across replicates. In addition, many strains in this collection formed filigreed mats without a spoke/lobe structure (Appendix A.1) and would not be quantifiable by those metrics.

The quantitative values across the flocculation, invasion, and polystyrene adhesion assays are shown in Figure 2.3 contrasted with the qualitative image data. Across all three quantitative assays the strains with the strongest qualitative phenotypes (UWOPS87-242.1 for settling, 322134S for invasion, and SK1 and 322134S for adhesion) all correspond to the highest recorded quantitative measurements. Distinctions between strains with weak phenotypes are less clearly represented by the quantitative data but trends of increasing phenotypic strength are consistent across all three quantitative assays. The results of the quantitative assays compared across

biological replicates also demonstrate that these assays are a consistent, repeatable method for examining these phenotypes (see Figure 2.2).

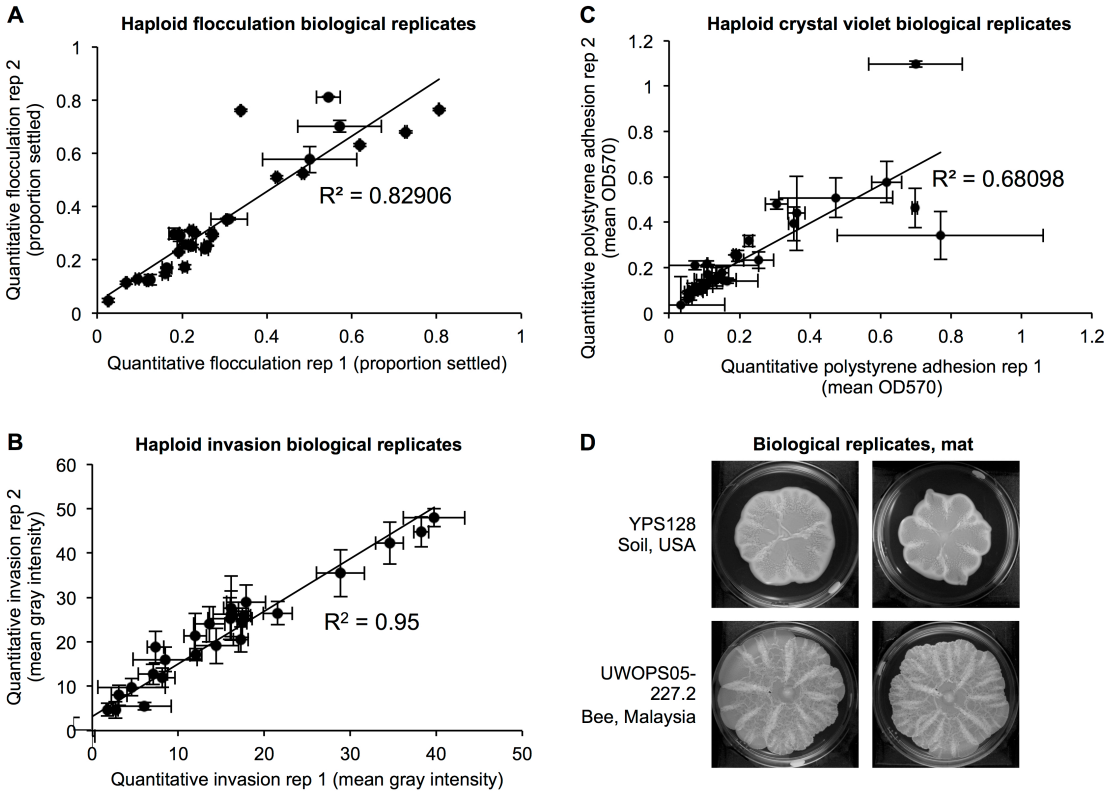


Figure 2.2. Haploid biological replicates demonstrate consistency of quantitative assays ***a*** Mean quantitative haploid flocculation data for one biological replicate (x-axis) is plotted against quantitative haploid flocculation data for a second biological replicate (y-axis) for correlation. Average taken and error bars calculated over three measurement replicates per strain. ***b*** Mean quantitative haploid invasion data for one biological replicate (x-axis) is plotted against quantitative haploid invasion data for a second biological replicate (y-axis) for correlation. Average taken and error bars calculated over three technical replicates per strain. ***c*** Mean quantitative haploid polystyrene adhesion data for one biological replicate (x-axis) is plotted against quantitative haploid polystyrene adhesion data for a second biological replicate (y-axis) for correlation. Average taken and error bars calculated over two technical replicates per strain. Outlier A shows the measurements from strain 322134S, and outlier B the measurements from strain SK1. Each of these strains was highly flocculent, and as a result the biofilm formed on polystyrene stained inconsistently across biological replicates and, in some cases, across technical replicates as well. If these outliers are eliminated from the analysis, the R^2 value becomes 0.83. ***d*** Biological replicates of complex mat formation for two biological replicates of representative haploid strains. Photos taken at day 13 of growth.

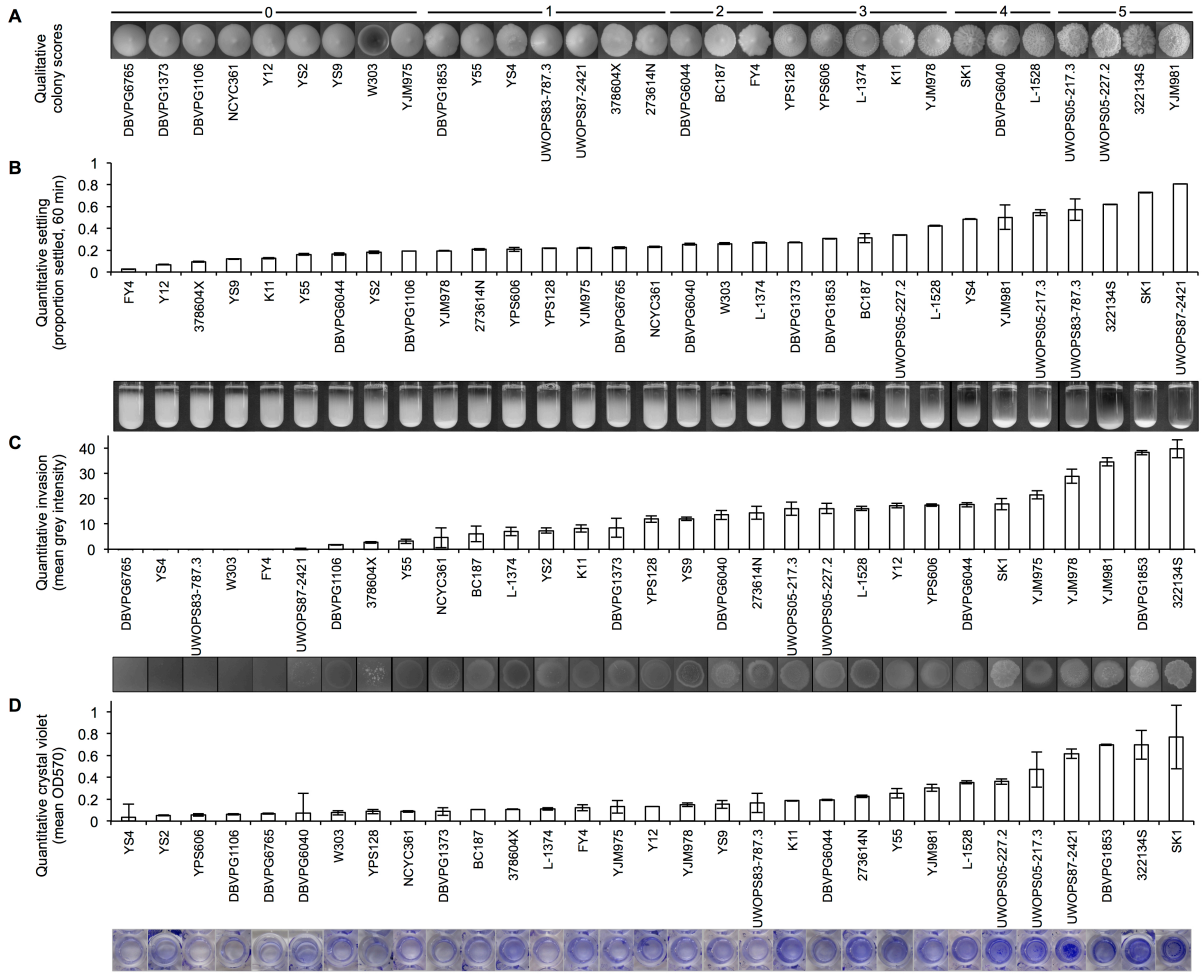


Figure 2.3. Quantitative measures of biofilm-related phenotypes accurately represent qualitative data. Haploid strains on the x-axis are ranked by increasing phenotypic strength. For procedural details of each method of quantification, see Materials and Methods. ***a*** Qualitative colony classifications at day 13 of growth, from (0) weak to (5) very strong. ***b*** Quantitative flocculation measured as the percent of culture settled at 60 minutes. Average of three measurements per image. ***c*** Quantitative invasion data measured as the mean grey intensity of the invaded spot. Average of three technical replicates. ***d*** Quantitative crystal violet data measured as OD570 absorbance of crystal-violet stained biofilms. Average of two technical replicates.

2.2.3 *Biofilm-related phenotypes show complex correlations*

Our results demonstrate that the phenotypic strength of a strain in one assay is not necessarily predictive of its strength across another. Of the six SGRP strains shown in Figure 2.1, only two (UWOPS05-227.2 and 322134S) have strong phenotypes across all assays. We were interested

in determining if any of these phenotypes show correlations with each other and to what extent each additional assay is providing new information about the behavior of the strain.

For the two assays for which only qualitative data were available (complex colony morphology and complex mat formation) we performed a Kendall's tau rank correlation test on the qualitative scores across all strains for a single biological replicate in a pairwise fashion. At a significance threshold of $p<0.05$, we found that these two phenotypes were highly correlated, with a p-value of 1.03×10^{-6}, an expected result given that the only difference between these assays is the agar concentration in the media. For assays with quantitative data, we binned the data across the associated qualitative colony scores (Figure 2.4 A, B, and C) to look for trends associated with increasing colony morphology complexity. Flocculation (Figure 2.4A) showed no correlation with colony morphology by this method, but there was a trend associating increasing invasive ability (Figure 2.4B) and polystyrene adhesion (Figure 2.4C) with increasing complexity in colony morphology.

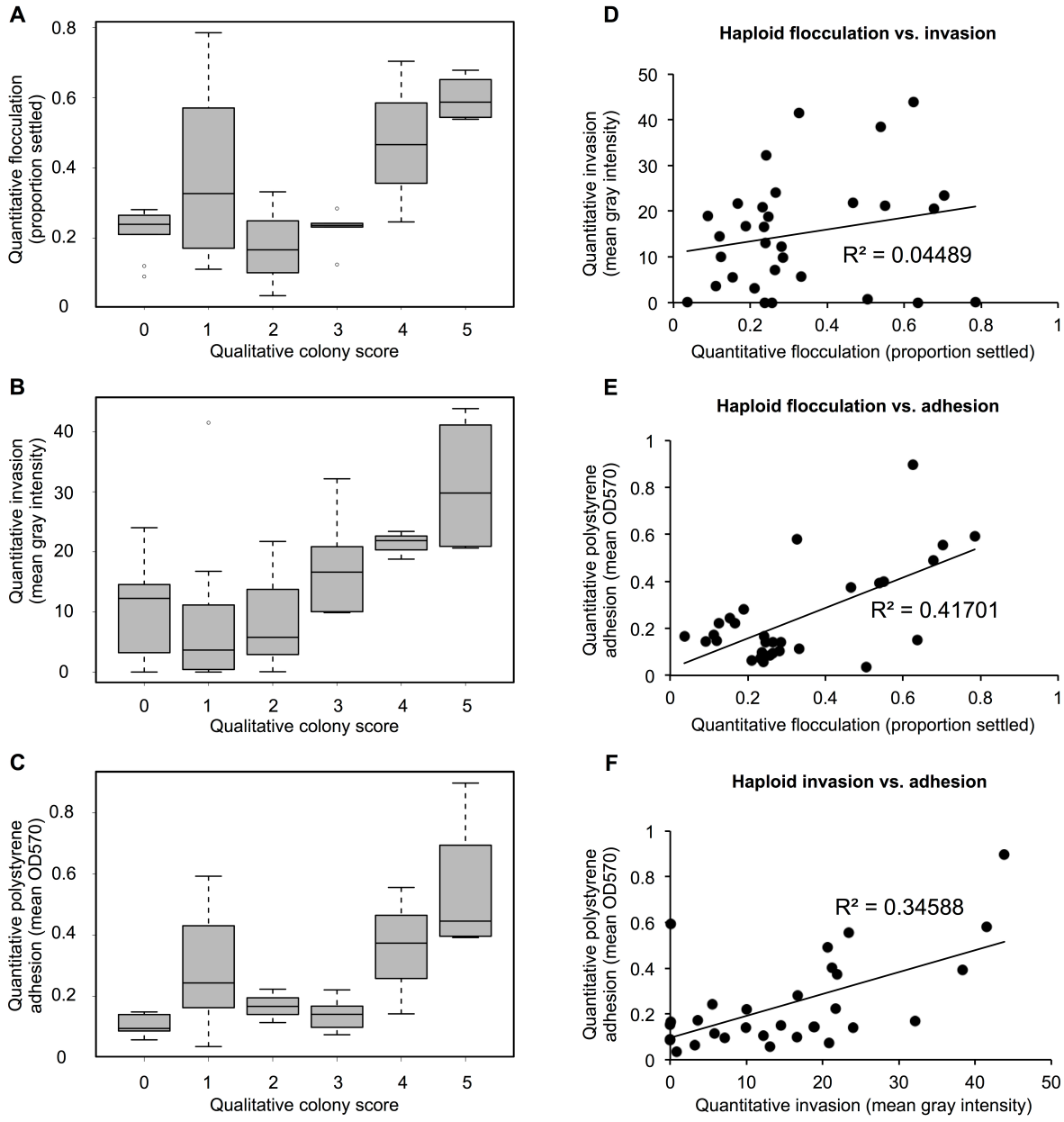


Figure 2.4. Some biofilm-related phenotypes are correlated. All data is for 31 haploid strains. Qualitative data shown is for a single haploid replicate; quantitative data is averaged across two biological replicates. ***a-c*** Mean quantitative data for (***a***) flocculation, (***b***) invasion, and (***c***) adhesion is shown on the y-axis binned according to haploid colony complexity scores (x-axis) ***d-f*** Mean quantitative data is plotted for correlation, with (***d***) flocculation (x) against invasion (y), (***e***) flocculation (x) against adhesion (y), and (***f***) invasion (x) against adhesion (y).

For the three assays with quantitative data available, we correlated the average quantitative values across technical replicates between each of the phenotypes. Despite the shared genetic

networks implicated in regulation of these traits, the majority of phenotypic correlations we observed were weak. We found no correlation between haploid flocculation and invasion phenotypes ($R^2 = 0.04$, Figure 2.4D) and a weak correlation between haploid invasion and polystyrene adhesion ($R^2 = 0.35$, Figure 2.4F). We found a stronger correlation between haploid flocculation and polystyrene adhesion ($R^2 = 0.42$, Figure 2.4E) though this correlation is still weaker than those between biological replicates ($R^2 = 0.83$ and $R^2 = 0.68$ from flocculation and adhesion biological replicates, respectively (Figure 2.2B and D)).

It is noteworthy that these pairwise comparisons of phenotypes fail to include information on the genotype of these strains, and it is possible that some of the weak correlations observed between phenotypes are not representative of true correlations but rather of genetic similarity between the strains being compared. In order to begin to address this possibility, we examined the distribution of phenotypes across the phylogeny for these strains (Figure 2.5A), generated in the Yeast Resource Center browser (http://www.yeastrc.org/g2p/). Filled bubbles on each tree represent the phenotypic scores, with a color scale based on the binned quantitative values of the phenotypes. There are examples of genetically similar strains with matching phenotypes (e.g., YPS606 and YPS128, UWOPS05-217.3 and UWOPS05-227.2) but also genetically similar pairs with very different phenotypic signatures (e.g., DBVPG6765 and L-1374).

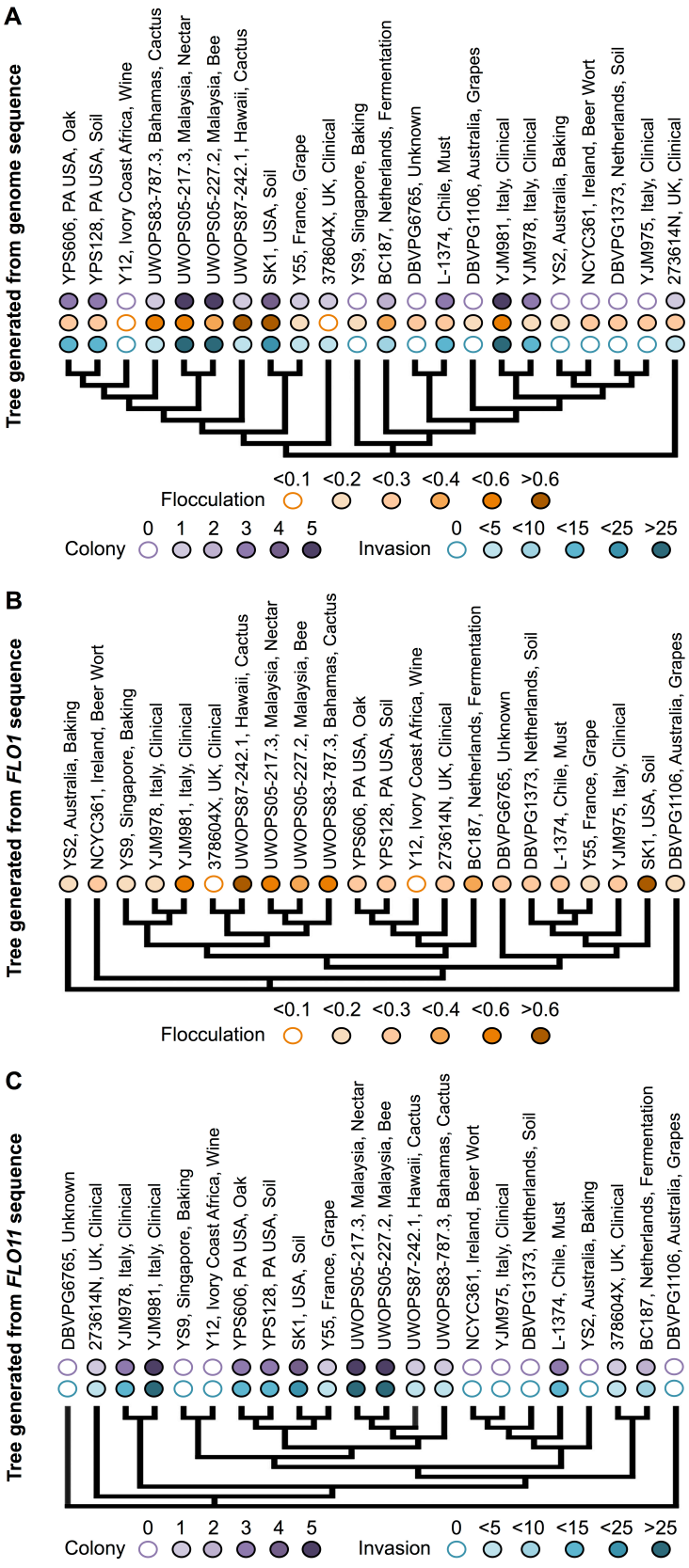


Figure 2.5. Genetic similarity is not the only predictor of phenotype.

a 22 SGRP strains clustered by genome sequence in the YRC phenome browser (http://www.yeastrc.org/g2p/). Qualitative and quantitative phenotypic data is shown above each branch, with increasing color intensity corresponding to increasing score. Qualitative colony data is scored and colored from intensity 0 to 5. Quantitative data was binned into 5 corresponding color intensities.

b Tree generated from *FLO1* gene sequence. Quantitative flocculation data is shown above each branch.

c Tree generated from *FLO11* gene sequence. Qualitative colony morphology and quantitative invasion data is shown above each branch.

Because the genome-wide phylogeny of these strains fails to account for the different evolutionary histories of individual genome regions, we also examined these traits at the level of single genes of interest. There are dozens of candidate genes that have been identified for their involvement in biofilm-related phenotypes in QTL studies (Brauer et al. 2006; Wilkening et al. 2014) and deletion collection studies (Granek and Magwene 2010; Ryan et al. 2012). From these identified candidate genes we focused on *FLO1* and *FLO11*, since they have also been extensively shown experimentally to be involved in flocculation (Miki et al. 1982; Smukalla et al. 2008) and cell-surface interactions (Lo and Dranginis 1998; W. S. Lo and Dranginis 1996; Váchová et al. 2011), respectively. The gene trees were generated in the Yeast Resource Center phenome project browser using ORF sequences and represent 22 resequenced strains from the collection. In Figure 2.5B, six of the most flocculent strains are shown juxtaposed on the *FLO1* gene tree; four cluster together while two strains, YJM981 and SK1, are genetically divergent at that locus. This indicates a possible role for certain very similar alleles of *FLO1* in the flocculation phenotype, with YJM981 and SK1 as possible points for comparison. YJM981 also has a different phenotype from its most genetically similar strain at the *FLO1* locus, YJM978. The relationship between *FLO11* genetic variation and phenotype is stronger (Figure 2.5C). Two phenotypes are shown here, complex colony morphology and invasion, both known to be associated with *FLO11* expression (Galitski et al. 1999; Granek and Magwene 2010). There are several instances of genetically similar pairs with identical or similar phenotypes (YS9 and Y12, YPS606 and YPS128, UWOPS05-217.3 and UWOPS05-227.2, NCYC361 and YJM975, 376804X and BC187), and instances of strains similarly related genetically at this locus with different phenotypes show weak differences only (SK1 and Y55, YJM978 and

YJM981). To effectively understand the role of genetic variation in these biofilm-related phenotypes a more targeted approach might be successful.

2.2.4 *Phenotypic pattern does not cluster according to environmental or geographical niche*

In order to determine if any of the phenotypic fingerprints of these strains clustered by their niche of origin, we examined the hierarchical clustering of the qualitative and quantitative scores by Euclidean distance using the simplified designations for each strain provided by Liti and collaborators (Liti et al. 2009). In Figure 2.6, two primary clusters emerge with strong phenotypes, with the cluster including strains DBVPG6044 through YJM978 driven by strong colony morphology and mat complexity, and the cluster including strains SK1 through YJM981 with strong phenotypes across all assays. No strong clusters emerge for the rest of the strains with weaker phenotypes. Some strains with similar origin niches do cluster together, including Pennsylvania strains YPS128 and YPS606 in the top cluster, and Malaysian strains UWOPS05-217.3 and UWOPS05-227.2 in the second cluster. Both of these pairs also have a high degree of sequence similarity, shown through their relatedness at both the *FLO1* and *FLO11* loci and genome wide in Figure 2.5. However, there are also several exceptions where strains with very similar niches do not cluster together. Examples include the five clinical strains, cactus strains UWOPS87-242.1 and UWOPS83-787.3, and Chilean must strains L-1374 and L-1528, which do not cluster together phenotypically despite sharing both the same ecological and geographical niche.

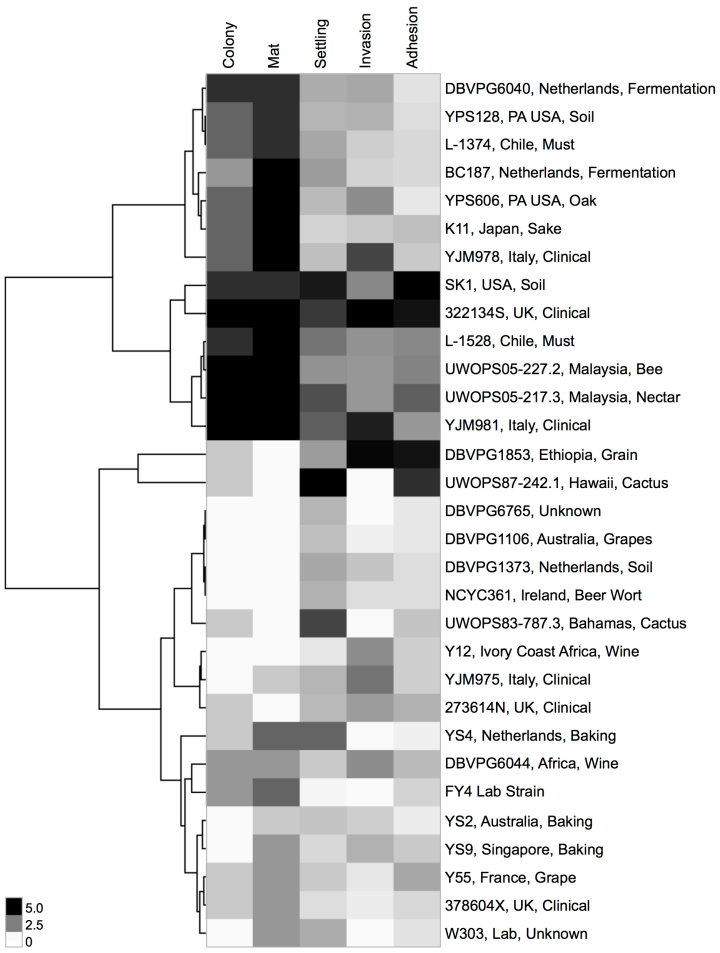


Figure 2.6. Few strains cluster phenotypically according to niche of origin. Hierarchical clustering of qualitative phenotypes (scores 0 to 5) and quantitative phenotypes (linear normalization to scale 0 to 5) for 31 haploid strains with simplified niche labels according to Liti et al (Liti *et al*. 2009). Cluster 3.0 (de Hoon *et al*. 2004) generated dendrogram based on Euclidean distance between strains; clustering visualized in JavaTreeView (Saldanha 2004). No phenotype (0) registers as white; very strong phenotype (5) registers as black.

2.2.5 *Colony morphology and agar invasion phenotypes in some natural isolates might be driven by prion content*

We hypothesized that non-genetic causes may affect some of the correlations and lack thereof that we observed. Previous work on natural isolates has shown that up to 8% of natural isolates may harbor prions (Halfmann et al. 2012). Some of the phenotypic changes in natural isolates cured for prions include differences in colony morphology, invasion, and flocculation phenotypes (Holmes et al. 2013). To determine if the biofilm-related phenotypes in these natural isolates were driven by prions, we cured the strains of prions according to the procedure outlined

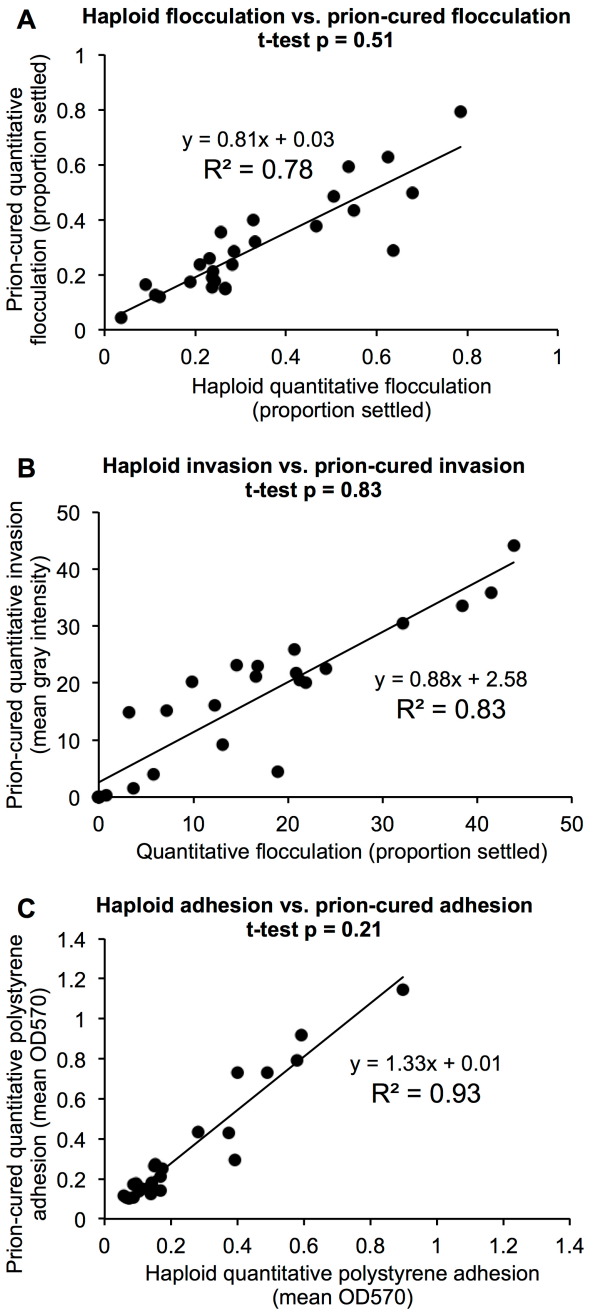


Figure 2.7. Prion-cured vs. haploid correlation plots show reproducibility ***a*** Mean quantitative haploid flocculation data for all biological replicates (x-axis) is plotted against quantitative prion-cured flocculation data across all cured replicates (y-axis) for correlation. P-value is shown for two-tailed t-test assuming unequal variance. ***b*** Mean quantitative haploid invasion data for all biological replicates (x-axis) is plotted against quantitative prion-cured invasion data across all cured replicates (y-axis) for correlation ***c*** Mean quantitative haploid polystyrene adhesion data for all biological replicates (x-axis) is plotted against quantitative prion-cured polystyrene adhesion data across all cured replicates (y-axis) for correlation.

in Holmes et al (Holmes et al. 2013) (see Materials and Methods). We tested strains after passaging for respiration competence on glycerol plates and six strains repeatedly failed to be respiration competent after curing. For the remaining 25 cured strains we completed a full phenotypic panel across four assays (Appendix A.2), excluding complex mat formation due to its strong correlation with complex colony morphology. We observed no significant phenotypic

differences from the haploid strains in our cured strains across all phenotypes except colony morphology and agar invasion. For nine of those strains we also generated an additional, separately cured replicate to confirm the cured phenotype; these replicates are shown in Appendix A.2 as well.

Three strains (273614N, YJM978, and YPS128) showed a reduction or variation in complex colony morphology after curing. Two additional cured strains (DBVPG1853 and W303) exhibited changes in colony morphology that were not carried through all cured replicates. All cured phenotypes are shown in Appendix A.2. These five strains had little shared history, spanning four ecological and five geographical niches. The agar invasion phenotype has both a characteristic intensity and a characteristic invasion pattern for each strain that is highly reproducible across biological replicates (Appendix A.1). For strains YPS128, DBVPG1106, Y12, YS2, YS4, and YS9 the prion-cured strains exhibited changes in both invasion intensity and pattern; Y12 also demonstrated differences between cured replicates. These differences are reflected in the correlation plot for haploid versus prion-cured strains across the invasion phenotype (Figure 2.7B) with an R^2 value of 0.83, lower than the haploid biological replicates ($R^2 = 0.95$). The six strains with altered invasion phenotypes in their prion-cured versions fall across four ecological niches and five geographical niches; all of the baker strains (YS2, YS4, YS9) and all of the Australian strains (DBVPG1106, YS2) in the collection potentially harbor prions.

2.2.6 *Diploid strains show changes in correlations between biofilm-related phenotypes*

In order to determine how ploidy may affect these traits, we completed a full phenotypic panel of all six assays, including complex mat formation, on the homothallic diploid SGRP collection. The complete panel is shown in Appendix A.3. We also examined an additional

phenotype, filamentous growth, which is specific to the diploid strains (Gimeno et al. 1992; Liu, Styles, and Fink 1996). We binned the quantitative invasive growth scores into five groups according to the same metric used in Figure 2.5 and compared those scores to the qualitative filamentous growth scores with a Kendall's tau rank correlation test. These two phenotypes are significantly correlated, with a *p*-value of 0.04. We repeated the low-nitrogen filamentous growth assay for the haploid collection as a control to ensure that none of the haploid strains demonstrated a filamentous growth phenotype. We observed no filamentous growth for any haploid strains (data not shown).

We repeated the correlation analyses we performed on the diploid collection to observe how the relationships between traits change in the diploids. We compared the quantitative diploid data across the three quantitative traits (flocculation, invasion, adhesion) with the complex colony morphology scores and observed a weak trend of increasing colony morphology associated with increased invasion (Figure 2.8B), similar to the haploids. We also performed a Kendall's tau rank correlation test on the diploid qualitative data; we found that the relationship between the colony morphology and mat formation assays changed with the change in ploidy, with the previous correlation lost (new p=0.08, above the significance threshold p<0.05). Several other correlations are lost in the diploid data, shown in Figure 2.8. There is no correlation between any of the quantitative phenotypes (Figure 2.8D, E, and F) with the correlation between flocculation and adhesion we observed in the haploids specifically lost in the diploids (R^2 = 0.005, Figure 2.8E).

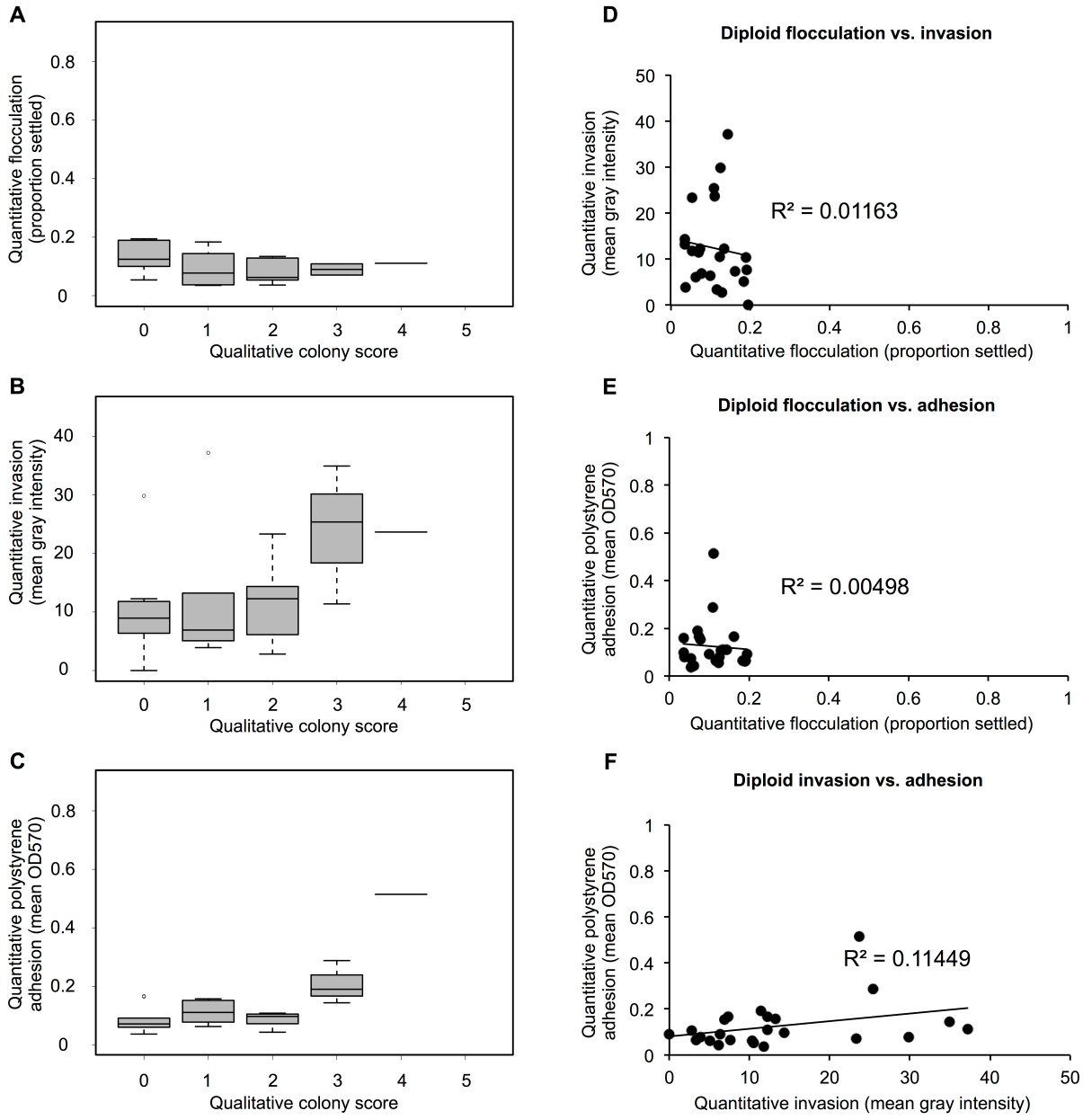


Figure 2.8. Diploid strains show weaker correlations between biofilm-related phenotypes. All data is for 24 diploid strains and averaged across technical replicates (see Materials and Methods). ***a-c*** Mean quantitative data for (***a***) flocculation, (***b***) invasion, and (***c***) adhesion is shown on the y-axis binned according to diploid colony complexity scores (x-axis) ***d-f*** Mean quantitative data is plotted for correlation, with (***d***) flocculation (x) against invasion (y), (***e***) flocculation (x) against adhesion (y), and (***f***) invasion (x) against adhesion (y).

Diploid strains do not uniformly have reduced phenotypic strength with respect to haploid strains

The strength and complexity of several of these biofilm-related phenotypes have been reported to be influenced by ploidy. Galitski and Fink and Reynolds and Fink observed that an increase in ploidy in laboratory strain Σ1278B results in both decreased invasive ability on agar as well as decreased mat complexity (Galitski et al. 1999; Reynolds and Fink 2001). Granek et al also reported a correlation between increased ploidy and decreased complex colony morphology across natural isolates (Granek and Magwene 2010).

Across the five primary phenotypic assays we observed that the strength of the phenotypes differed with respect to the haploid strains, and were generally attenuated across all phenotypes (Figure 2.9). No diploid strains achieved a complex colony morphology score of 5 (Figure 2.8A, B, and C), a complex mat score of 5, or a flocculation score of higher than 0.3, where the highest haploid flocculation score was 0.8 (strain UWOPS83-787.3) and 16 strains achieved a score of 0.3 or higher. These trends are reflected in the haploid/diploid correlation plots in Figure 2.9.

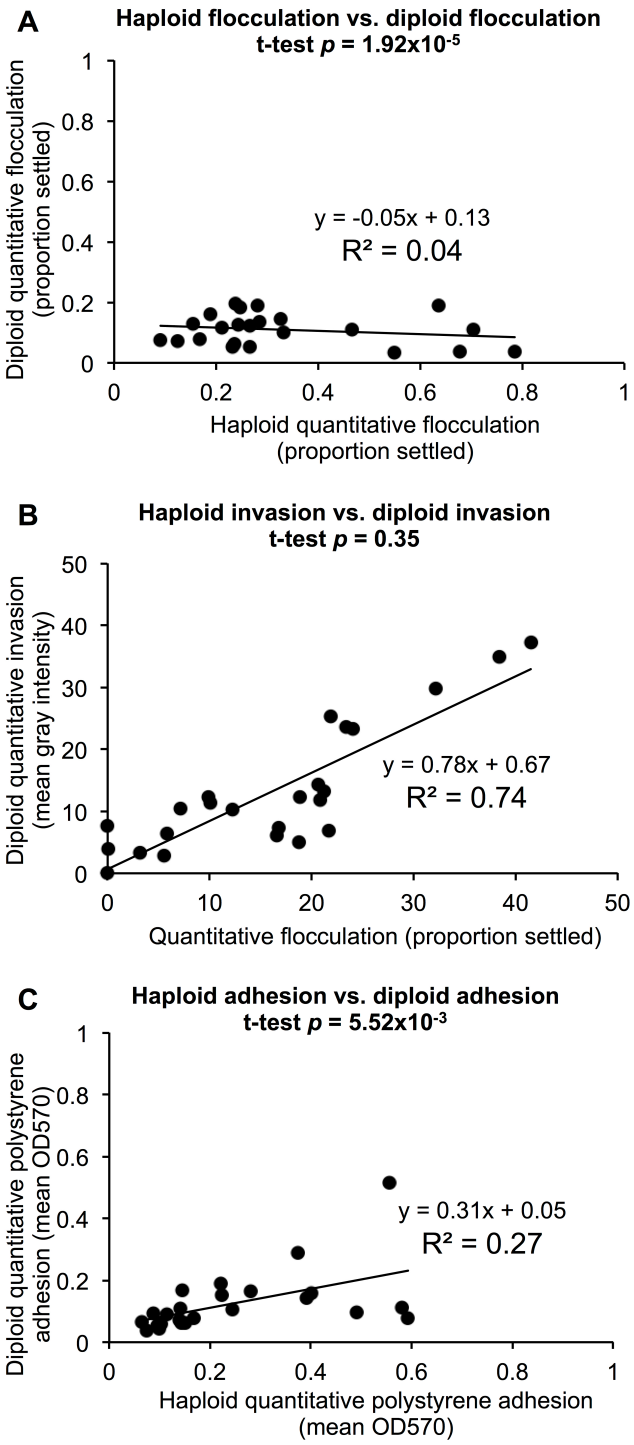


Figure 2.9. Diploid quantitative phenotypes are weaker than haploid quantitative phenotypes except for invasion. ***a*** Mean quantitative haploid flocculation data for two biological replicates (x-axis) is plotted against quantitative diploid flocculation data (y-axis) for correlation ***b*** Mean quantitative haploid invasion data for two biological replicates (x-axis) is plotted against quantitative diploid invasion data (y-axis) for correlation ***c*** Mean quantitative haploid polystyrene adhesion data for two biological replicates (x-axis) is plotted against quantitative diploid polystyrene adhesion data (y-axis) for correlation.

In Figures 2.9A and 2.9C, the low slopes (m=-0.05 and m=0.31, respectively) and correlation coefficients (R^2=0.04 and R^2=0.27) demonstrate that in the flocculation and polystyrene adhesion assays, diploid strains have very different phenotypes from their haploid counterparts and these

phenotypes are weaker than the haploid phenotypes. These numbers are particularly striking compared to the haploid/prion-cured differences, which should be negligible if the phenotypes are not strongly influenced by prions. In fact we do find that to be the case with 20 out of 26, 20 out of 23, and 23 out of 25 prion-cured strains showing no significant change in flocculation, invasion, and adhesion, respectively, compared to the haploids (Table 2.4).

Table 2.4 Quantitative changes in prion-cured strains vs. haploid strains

	FLOCCULATION[a]	INVASION[b]	ADHESION[c]
Significant decrease by FDR correction*	5	1	0
No change by FDR correction	20	20	23
Significant increase by FDR correction	1	2	2

The numbers of prion-cured strains that showed differences when compared to the original haploid strains across three quantitative metrics are shown. Quantitative values were compared using a two-tailed t-test assuming unequal variance across all technical replicates for each strain. *P*-values were evaluated for significance using the *q*-values package from Storey with the Benjamini-Hochberg method (Storey 2002).

* FDR = 0.05

[a] 26 strains

[b] 23 strains

[c] 25 strains

These values confirm that the magnitude of the changes in the diploid strains is well outside the range of biological noise in the assay. The invasion assay correlation, shown in Figure 2.9B, shows a different pattern. The high R^2 value of 0.74 is slightly lower than the biological replicate noise level (R^2 of 0.95 for invasion, Figure S1 C) and the slope m=0.77 close to 1 indicates that the haploid and diploid quantitative invasion values are very similar. Indeed,

contrary to predictions from literature, we observe two cases of increases in invasive ability in the diploid strains, shown in Figure 2.8 and Table 2.5.

Table 2.5 Quantitative changes in diploid strains vs. haploid strains

	FLOCCULATION[a]	INVASION[b]	ADHESION[c]
Significant decrease by FDR correction*	22	6	11
No change by FDR correction	1	15	13
Significant increase by FDR correction	0	2	0

	COLONY	MAT
Decrease by qualitative score	13	12
No change by qualitative score	9	3
Increase by qualitative score	2	9

The numbers of diploid strains that showed differences when compared to the related haploid strains across three quantitative metrics are shown. Quantitative values were compared using a paired two-tailed t-test assuming unequal variance across all technical replicates for each strain. *P*-values were evaluated for significance using the *q*-values package from Storey with the Benjamini-Hochberg method (Storey 2002).

* FDR = 0.05

[a] 23 strains

[b] 23 strains

[c] 24 strains

This increase in invasive ability was statistically significant in a *q*-value test between average haploid and diploid values for strains UWOPS87-242.1 and UWOPS83-787.3. This increase in invasive ability is also visible in the invasion images (Figure 2.10).

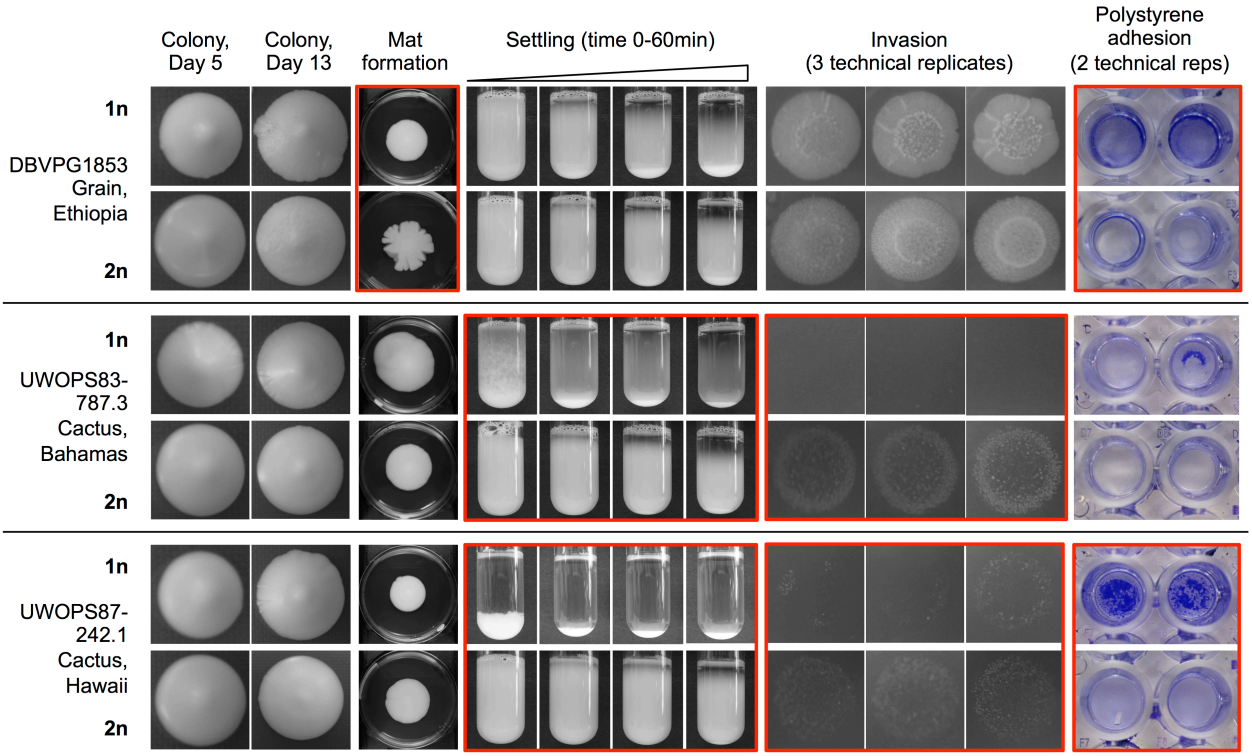


Figure 2.9. Examples of phenotypic differences depending on ploidy Three representative strains are listed with their formal name and origin and are shown across five different phenotypes, with ploidy designated "1n" or "2n" for haploid and diploid strains. Red boxes indicate notable changes in phenotype between haploid and diploid versions of each representative strain. Image details and experimental methods are the same as Figure 2.1.

Though we saw examples of diploid strains with weakened phenotypes across all metrics, there were several notable counterexamples. In Figure 2.10, weakened phenotypes are shown for strains DBVPG1853 and UWOPS87-242.1, which lose their polystyrene adhesive ability as diploids, and strains UWOPS83-787.3 and UWOPS87-242.1, which both lose their strong flocculation ability as diploids. However, some exceptions emerge: strain DBVPG1853 gains a complex mat as a diploid, and invasive ability improves in the diploid versions of UWOPS83-787.3 and UWOPS87-242.1, especially for UWOPS83-787.3.

Across qualitative scores for a single haploid biological replicate, nine strains showed a more complex mat as diploids than as haploids, though in five of these cases the qualitative scores show a weak gain of complexity only (Table 2.3). The exceptions in the quantitative

assays are also highlighted in a statistical comparison of the haploid and diploid data shown in Table 2.5, which shows the number of diploid strains that showed a significant increase or decrease across each phenotypic category with respect to the haploid strains (see Materials and Methods). With those conditions applied, 22 of 23 tested strains had a significant decrease in flocculation as diploids, but only six strains had a significant decrease in invasive ability. The most common scenario was to see no significant difference between haploids and diploids (15 out of 23 strains for invasion, 13 out of 24 strains for adhesion). Instances of no significant difference were not dominated by strains with weak phenotypes; of the 15 strains with no significant difference in invasive ability, only six strains had diploid mean gray intensity values less than 10.

2.3 DISCUSSION

Understanding connections between genotype and phenotype is one of the primary driving motivations behind biology and genomics today, yet many of the tools we are using to answer this question are constrained by lack of phenotypic diversity. Biofilm-related phenotypes in yeast, for example, are limited in laboratory strains and only now with natural yeast isolates are we able to examine the full diversity of phenotypes and their possible genetic origins. In this study we have demonstrated the incredible amount of phenotypic diversity present across biofilm-related phenotypes in just 30 natural isolates, a limited panel. With this phenotypic panel we have shown that despite the predicted involvement of similar genetic pathways, many of these biofilm-related phenotypes are uncorrelated or weakly correlated with each other and are potentially controlled by different genetic networks. We have also demonstrated that the majority of phenotypes we observe are not due to prions and are variable depending on ploidy, but not in predictable ways. General conclusions about the effects of ploidy on biofilm-related

phenotypes have been drawn in the literature based on only a few strain backgrounds and although these trends are generally observed in our data, there are also notable exceptions. These findings demonstrate the importance of using natural isolates to understand genotype-phenotype relationships because many predictions made in laboratory strains do not hold true for natural isolates.

In comparing complex colony morphology, complex mat formation, flocculation/settling, agar invasion, and polystyrene adhesion/crystal violet phenotypes in haploid natural isolates, we found two strongly correlated relationships (complex colony morphology and complex mat formation; flocculation and polystyrene adhesion) and two weakly correlated relationships (complex colony morphology and invasion; invasion and polystyrene adhesion). Given that most of the relationships we observed in Figure 2.4 were weakly correlated or uncorrelated, this suggests complex genetic interactions underlying each phenotype, though some of the stronger correlations could be explained by shared genetic drivers. All of these traits have been shown in previous work to be dependent on *FLO11* activity (Lo and Dranginis 1998; Reynolds and Fink 2001; Granek et al. 2013), including flocculation under specific conditions (Bayly et al. 2005) though flocculation is primarily considered to be a *FLO1*-mediated trait (Smukalla et al. 2008). This difference in genetic control is potentially responsible for the lack of correlation we observed between quantitative flocculation and invasion values (R^2=0.03), which is contrary to the correlation between these phenotypes observed in *S. paradoxus* (Roop and Brem 2013). The caveat to this relationship is that we observed that clumpy strains are less likely to form a consistent monolayer of cells in the polystyrene adhesion assay, instead retaining some of their clumpiness in the biofilm. These clumps absorb more of the crystal violet stain, potentially skewing the adhesion results in the direction of particularly flocculent strains. This could

potentially be addressed in the future by treating the cultures with a deflocculation buffer prior to either biofilm formation or staining. More interesting is the weak correlation between the invasion phenotype and the polystyrene sticking phenotype (R^2=0.29), as previous studies have shown that *FLO11* is involved in the manifestation of both phenotypes. This suggests that though these phenotypes have a connection to *FLO11* activity they are likely regulated by more complex genetic interactions.

Ryan and colleagues also recently examined several of these phenotypes in the *S. cerevisiae* Σ1278B knockout collection (Ryan et al. 2012), finding that with a few exceptions different "growth programs" like filamentous growth and mat formation were controlled by many different genes in an only semi-overlapping fashion. Given that we observed only one very strong correlation between biofilm-related phenotypes (complex colony morphology and complex mat formation) this compartmentalized management of growth programs is what our data suggest as well. The practical outcome of these findings is that to consider the biofilm-related phenotype of one of these natural isolate strains fully, many different assays are needed to accurately describe the characteristics of the strain. The independence of these phenotypes further provides future opportunities to differentiate and classify the genes and pathways affecting each biofilm-related phenotype.

We hypothesized that diploid phenotypes would be weaker in general than the haploid phenotypes for the corresponding strains, according to the work of Galitski and Fink, Reynolds and Fink, and Granek and Magwene, who each showed that strength of invasion, strength of complex mat formation, and strength of complex colony morphology decrease with ploidy (Galitski et al. 1999; Reynolds and Fink 2001; Granek and Magwene 2010) and, in the case of Galitski *et al*, that this ploidy-regulated difference was independent of hetero- or homozygosity

at the mating type locus. We observed this same general trend across all phenotypes with some interesting exceptions; across natural isolates, the relationship between the haploid and diploid phenotypes varied according to strain and phenotype examined. Within our representative set of strains in Figure 2.10, there are examples of phenotypes both stronger and weaker in the diploid strains, which is a characteristic of the collection as a whole (see haploid strains in Appendix A.1 and diploid strains in Appendix A.3). Most importantly, we have recorded examples in this dataset of significantly increased phenotypic strength in the diploid strains across the complex mat formation and invasion assays, both assays in which an attenuated diploid phenotype was predicted based on experiments in laboratory strains. Our finding that predictions about the effect of ploidy on biofilm-related phenotypes are not necessarily borne out in natural isolates is consistent with the findings of Zörgö and colleagues, who examined ploidy-environment interactions for SGRP haploid and diploid natural isolates and discovered countless exceptions to predictions of how ploidy would impact a strain's response to environmental perturbations (Zörgö et al. 2013). Other studies with this collection suggest laboratory strain S288C is a phenotypic outlier with respect to natural isolates (Warringer et al. 2011), a further reminder that the phenotypes recorded from laboratory strains of *S. cerevisiae* are not necessarily generalizable to natural isolates. The SGRP collection and other collections of natural isolates will allow us to test predictions generated from analyzing laboratory strains and see under what conditions those predictions hold true.

Given a larger subset of strains, the conclusions we draw about the correlations between these phenotypes would undoubtedly be refined. With a sample set of more than 30 strains the data generated about phenotypic relationships and especially niche relationships would particularly benefit; for some of the niches in which we observed clustering in Figure 2.6 there

were only a few representatives from a given niche (e.g. only three African strains, only three baking strains, only two Malaysian strains). As a result the conclusions we draw come with a significant caveat. Though our clustering analysis is likely underpowered, the lack of correlation we saw between strength of biofilm-related traits and any particular niche is consistent with observations from Warringer et al (Warringer et al. 2011) who demonstrated that trait variation across an enormous panel of phenotypes was not driven primarily by the niche of origin.

With a larger collection it would also be critical to develop high-throughput and potentially automated versions of the standard phenotypic assays described in this work. We could potentially refine the picture further by expanding the phenotypic assays to examine natural isolates' behavior under nutrient-limited conditions, which have been shown to alter biofilm-related phenotypes (Palecek, Parikh, and Kron 2000; Granek and Magwene 2010). There is also room to improve the scope of the quantitative assays. Right now, the assays only capture one primary dimension of each phenotype but there is additional variation that stratifies the phenotypes even further. For example, in the flocculation assay plot profiles as shown in Appendix A.4 and A.5, there is additional information about the way a strain settles in the slope and integral of the plot profile line that is not captured by taking a single distance measurement. Similarly, in the invasion assay as shown ranked in Figure 2.3B, strains with very similar quantitative invasion scores have clearly different invasion structures, with the majority of the intensity driven by a strong edge (e.g. L-1528 and YS9), a punctate pattern (e.g. 378604X) or a completely invaded patch (e.g. Y12 and K11); this fine-scale information is lost in the current invasion quantification metric. There is also room to generate improved quantitative data across the two assays that remain qualitative in this study, particularly the complex colony morphology assay. Quantitative methods for measuring complex colony morphology are becoming available

using time-course imaging and specialized image analysis software (Ruusuvuori et al. 2014). In future work with natural isolates it would be useful to assign quantitative scores to the complex colonies in this natural isolate collection as well.

We anticipate that the phenotypic data generated in this study will be a valuable platform for examining genotype-phenotype questions in a targeted way. Now that we appreciate the complexity of biofilm-related phenotypes across these strains, we plan to examine the contributions of variation in individual alleles to each phenotype in the context of different genetic backgrounds. We also hope that the phenotypic data generated in this study will be a useful community resource for the many groups now working with strains from this collection. Ultimately, understanding the contributions of variation in known biofilm-related genes to each of these phenotypes could also contribute to effective engineering of strains with very specific biofilm-related traits.

2.4 MATERIALS AND METHODS

2.4.1 *Strains and media used in this study*

SGRP strains used in this study (Liti et al 2009) were purchased from the National Collection of Yeast Cultures (http://www.ncyc.co.uk/sgrp.html) and are listed in Table 2.1. Strain identity and ploidy was verified for a subset of diploid strains by sequencing, sporulation, and cytometry (data not shown). Standard growth medium for plates and liquid overnight cultures was yeast extract peptone dextrose (YPD) media, with 2% glucose and 2% agar for plates. Growth in the polystyrene adherence assay was done in synthetic complete media with 2% glucose or 0.1% glucose. Standard media were prepared according to Sherman (Sherman 2002). Diploid filamentous growth was assayed on Synthetic Low Ammonia Histidine Dextrose (SLAHD) media prepared according to a modified recipe from Lorenz and Heitman (Gimeno et al. 1992;

Lorenz and Heitman 1997) with 50 μM ammonium sulfate, 2% glucose, 2% bacto-agar, 1.7 g YNB without ammonium sulfate or amino acids, and 0.2mM histidine hydrochloride per liter. Prion curing was conducted on 3 mM guanidine hydrochloride plates prepared according to Holmes and Halfmann (Holmes et al 2013).

2.4.2 *Standardized phenotypic assay protocols*

Measuring complex colony morphology: Two biological replicates of strains were inoculated from colonies into 200 μL YPD liquid overnight cultures in 96-well plates. They were diluted to concentrations of 10^{-5}, 10^{-6}, and 10^{-7} for haploid strains, with the second biological replicate at concentration 10^{-6} only. 200μL of all four dilutions of culture was plated using glass beads on 2% agar, 2% glucose YPD plates. The colony plates were incubated at 30˚ for three days and at 25˚ for ten additional days. The colonies were photographed and scored for complexity on days 5, 9 and 13. All imaging in this and other assays was done using a Canon Powershot SD1200 IS digital camera. This procedure was maintained for all collections (haploid SGRP strains, diploid SGRP strains, and prion-cured strains) with modifications only in the dilutions plated (10^{-6} only for diploid and prion-cured strains). Colony complexity was scored from 0 (smooth colony) to 5 (very strong complex colony morphology) according to the metric described in Table S2, adapted from Granek et al (Granek and Magwene 2010).

Measuring complex mat formation: Strains were inoculated by toothpick from colonies onto the center of a 0.3% agar, 2% glucose YPD plate left to dry for one day. The mat formation plates were sealed with parafilm and incubated upright at 25˚ for 13 days (Reynolds and Fink 2001). On the final day, the complex mats were photographed and scored for complexity according to the metric described in Table S2. This assay was performed once across all strains in the diploid and prion-cured collections and for two biological replicates in the

haploid collection. A second biological replicate was done for eight of the nine diploid strains with phenotypes that differed from the haploid strains. One of these replicates with a weak phenotype was inconclusive by visual scoring.

Measuring flocculation as settling in liquid culture: The settling assay measures clumping and settling of cells to the bottom of a 2% glucose YPD liquid culture (Johnston and Reader 1983; Liu, Styles, and Fink 1996; Smukalla et al. 2008). Strains were inoculated from colonies into 5 mL YPD liquid cultures and grown for 20 hours at 30˚. The strains were then vortexed vigorously in the culture tubes and allowed to settle. For each strain, four photographic timepoints were taken at 0, 15, 30, and 60 minutes with the culture tubes undisturbed between time points. This assay was performed once across all strains in the diploid and prion-cured collections and for two biological replicates in the haploid collection.

Quantitative flocculation assay: Quantitative flocculation analysis was conducted on the images from the settled 60-minute time point for each strain across two biological replicates of the haploid collection, the diploid collection, and the prion-cured collection. The images were converted to black and white in image processing software Picasa version 3.9.16.37 and analyzed in image processing software ImageJ version 1.47 (Abramoff, Magalhães, and Ram 2004). A line was drawn from the meniscus to the bottom of the culture tube in each image and used to measure the maximum grey intensity and the plot profile of intensity along the line. From the plot profile data, the x-coordinate along the line at which half of the maximum grey was reached was calculated, and used to calculate a ratio: x coordinate of half of maximum grey intensity / length of the line. Three of these measurements were taken per image and the average ratio across plot profiles calculated to determine, for each strain, the percentage of the tube cleared at 60 minutes.

Measuring agar invasion: The agar invasion assay measures a strain's ability to invade solid agar media, quantified by the density of cells remaining in the media when colonies are washed off the surface. Strains were inoculated from colonies into 5 mL YPD liquid overnight cultures. They were grown to saturation and standardized to an OD between 0.5 and 0.8. The 2% agar, 2% glucose YPD invasion plates were spotted with all strains from a collection (haploid SGRP strains, diploid SGRP strains, or prion-cured strains) and the FY4 (s288C) reference strain with a 1 μL spot per strain. Four invasion plates were inoculated per collection to provide technical replicates. Invasion plates were grown at 30˚ for five days after which time the colonies were washed off the surface of the plates under a stream of distilled water using a gloved finger (Liu, Styles, and Fink 1996; Roop and Brem 2013). The remaining invaded spots were photographed for each spot on each of the four plates. The washed plates were incubated again at 30˚ for 24 hours, then washed and photographed again to observe additional growth from cells trapped in the agar (Drees et al. 2005; Ryan et al. 2012). This assay was performed once across all strains in the diploid and prion-cured collections and for two biological replicates in the haploid collection. Eight diploid strains had a statistically significant difference in invasive ability compared to the haploid strains and these strains were selected for a second biological replicate.

Quantitative invasion calculations: From the four technical replicates of each strain, the three best images were selected for quantitative analysis. The images were converted to black and white in image processing software Picasa version 3.9.16.37 and analyzed in image processing software ImageJ version 1.47. To quantify the invasion for each spot, the mean grey intensity of the spot and surrounding background was calculated in ImageJ. The mean grey intensity of the entire image was also calculated in ImageJ and used to subtract an average

background value from the invaded spot. The formula used for background correction calculated the background-corrected intensity value as $[(A_t*I_t)-(A_s*I_s))/(A_t-A_s)]$ where A_t is the area of the total image, A_s is the area of the invaded spot, I_t is the mean grey intensity of the total image, and I_s is the mean grey intensity of the invaded spot. For any strains for which background correction yielded a negative invasion value, the value was corrected to zero. This analysis was performed for the full haploid, diploid, and prion-cured collections and for an additional biological replicate of the haploid collection.

Polystyrene adhesion with crystal violet assay: Polystyrene adhesion is measured by growing biofilms in flat-bottom polystyrene 96-well plates and staining for remaining cells using crystal violet dye. Strains were inoculated into SC cultures with 2% glucose and grown to an OD between 0.5 and 1.5 according to Reynolds and Fink. OD600 was measured on a BioTek Synergy H1 hybrid microplate reader. Cells were spun down and washed with distilled water then resuspended in SC media with 0.1% glucose to a standard OD of 1. 100 μL of culture was inoculated in two technical replicates, three for strains with enough culture volume after resuspension, into a 96-well Costar untreated flat-bottom polystyrene plate. The flat-bottom plate was incubated for 6 hours at 30˚. The cultures were then fixed and stained with 100μL per well of crystal violet dye solution (Fisher Scientific). Crystal violet dye was prepared as a 1% crystal violet solution in 100% ethanol and filter sterilized through a 0.2 μm filter. The crystal violet was incubated in the wells for 15 minutes at room temperature then decanted onto absorbent bench pads. The wells were washed three times with 300 μL sterile water, decanting to rinse. This assay was performed once across all strains in the diploid and prion-cured collections and for two biological replicates in the haploid collection.

Quantitative polystyrene adhesion with crystal violet assay: Stained biofilms were analyzed according to Reynolds and Fink. Biofilms were solubilized in 100 μL 10% SDS per well with a 30-minute incubation at room temperature. 100 μL sterile water was then added and 100 μL of solubilized dye was transferred from each well to a new flat-bottom 96-well plate. Absorbance was measured at 570nm, the absorbance of crystal violet, on a BioTek Synergy H1 hybrid microplate reader.

Diploid filamentous growth assay: Strains were plated following the complex colony morphology assay protocol described above onto Synthetic Low Ammonia Histidine Dextrose (SLAHD) plates, to encourage filamentous growth in a nitrogen poor environment. The colony plates were incubated at 30˚ for three days and at 25˚ for two additional days. The colonies were photographed and scored for filamentous growth under a brightfield microscope at between 20x and 50x magnification. The colonies grew at 25˚ for four additional days and were photographed and scored for additional filamentous growth on the ninth day of growth. Colonies were scored from 0 to 5 for filamentous growth according to the metric described in Table S2.

Curing strains of prions: To eliminate any prions in the natural isolates, all haploid strains were passaged four times on 3 mM guanidine hydrochloride YPD plates, bottlenecking through a single colony each time (Holmes et al. 2013). Clones were selected from the cured strains and passaged on 1.5% agar yeast-extract peptone glycerol media to check for respiration competence. Strains that were respiration competent were stored as prion negative strains in glycerol stocks; strains that were not respiration competent were excluded from further experiments. Prion-cured strains were phenotyped according to the same assay protocols described above. For seven strains that showed any potential phenotypic differences after curing

and two that did not, additional cured replicates were generated to verify the phenotypes and observe the effects of potential additional mutations resulting from the curing process.

2.4.3 *Correlations between quantitative phenotypes*

To determine whether any of the biofilm-related phenotypic traits were correlated, mean quantitative data across each phenotype was compared using a paired two-tailed t-test assuming unequal variance. For colony morphology and mat formation, the two qualitative phenotypes scored from 0 to 5, a Kendall's tau rank correlation test (Kendall 1938) was performed in R on the scores for each strain, comparing each phenotype pairwise.

2.4.4 *Hierarchical clustering of phenotypes by niche*

To determine whether any phenotypes clustered by niche of origin, the quantitative phenotypic values for each strain were linearly normalized to a scale of 0 to 5 and used as input into Cluster version 3.0 (Hoon et al. 2004). The strains were clustered hierarchically based on Euclidean distance between the quantitative phenotype vectors and visualized in Java Treeview (Saldanha 2004).

2.4.5 *Statistical comparison of haploid and diploid strains*

Quantitative values for haploid and diploid strains across the flocculation, invasion, and crystal violet assays were compared using a paired two-tailed t-test assuming unequal variance across all technical replicates for each strain. *P*-values were evaluated for significance using the *q*-values package from Storey with a False Discovery Rate of 0.05 and the Benjamini-Hochberg method (Storey 2002).

2.5 Acknowledgements

We thank Aimée Dudley for helpful comments on the manuscript; Celia Payen for editing and clustering advice; Rachel Gittelman, Maxwell Libbrecht, and Matthew Snyder for statistics consultations; Joshua Akey and Joseph Felsenstein for phylogeny advice; and Caitlin Connelly for information on the SGRP collection. This project was supported by grants R01 GM101091 and P41 GM103533 from the National Institute of General Medical Sciences from the National Institutes of Health. This material is based in part on work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454. E.H. was supported by T32 HG00035. M.D. is a Rita Allen Foundation Scholar and a Fellow in the Genetic Networks program at the Canadian Institute for Advanced Research.

Chapter 3. EXPERIMENTAL EVOLUTION REVEALS FAVORED ADAPTIVE ROUTES TO CELL AGGREGATION IN YEAST

The work described here was performed by Elyse Hope and collaborators. Sequencing analysis was conducted by Clara Amorosi with contributions from Caiti Smukowski Heil. Bulk Segregant Analysis and microscopy experiments were conducted by Kolena Dang with contributions from students participating in the Cold Spring Harbor Yeast courses 2014-2016. The 96 evolution experiments on which this work is based were performed by Aaron Miller. This chapter as written is pending review under this title and available as a pre-print on bioRxiv doi: 10.1101/091876

Yeast flocculation is a community-building cell aggregation trait that is an important mechanism of stress resistance and a useful phenotype for brewers; however, it is also a nuisance in many industrial processes, in clinical settings, and in the laboratory. Chemostat-based evolution experiments are impaired by inadvertent selection for aggregation, which we observe in 35% of populations. These populations provide a testing ground for understanding the breadth of genetic mechanisms *Saccharomyces cerevisiae* uses to flocculate, and which of those mechanisms provide the biggest adaptive advantages. In this study, we employed experimental evolution as a tool to ask whether one or many routes to flocculation are favored, and to engineer a strain with reduced flocculation potential. Using a combination of whole genome sequencing and bulk segregant analysis, we identified causal mutations in 23 independent clones that had evolved cell aggregation during hundreds of generations of chemostat growth. In 12 of those clones we identified a transposable element insertion in the promoter region of known flocculation gene

FLO1, and in an additional five clones we recovered loss-of-function mutations in transcriptional repressor *TUP1*, which regulates *FLO1* and other related genes. Other causal mutations were found in genes that have not been previously connected to flocculation. Evolving a *flo1* deletion strain revealed that this single deletion reduces flocculation occurrences to 3%, and demonstrated the efficacy of using experimental evolution as a tool to identify and eliminate the primary adaptive routes for undesirable traits.

3.1 INTRODUCTION

Experimental evolution is an essential tool for investigating adaptive walks, clonal dynamics, competition and fitness, and the genetic underpinnings of complex traits. One question experimental evolution enables us to explore is how often given the same conditions and selective pressures organisms will follow the same adaptive route (Gould's "tape of life") (Gould 1990; Orgogozo 2015). A primary platform for performing evolution experiments in the laboratory is the chemostat, a continuous culture device invented in 1950 by Monod (Monod 1950) and by Novick and Szilard (Novick and Szilard 1950). In a chemostat, new media is added and diluted at the same rate, maintaining constant growth conditions. Chemostat experiments have provided insight into the mechanisms of genome evolution and adaptation to a variety of selection pressures (reviewed in Gresham and Hong 2015). However, chemostats have been limited in their utility due in part to frequent selection for biofilms and cell aggregation, which have been observed since the advent of the chemostat and are thought to evolve due to selection by the physical constraints of the culture vessels. In 1964, Munson and Bridges recorded a selective advantage in a bacterial subpopulation that adhered to the wall of a continuous culture device (Munson and Bridges 1964). Topiwala and Hamer followed up on these findings in 1971 and suggested that encouraging this phenotype could actually lead to increased biomass output

(Topiwala and Hamer 1971), an idea that has enjoyed success in subsequent years: chemostats are such a successful system for growing biofilms that they are often used to grow biofilms intentionally by supplying additional substrates to encourage biofilm development (Poltak and Cooper 2011).

In the context of experiments concerning traits unrelated to wall growth and aggregation, however, the ease of biofilm evolution in chemostats represents a significant problem. Since wall growth and aggregation phenotypes develop as an adaptation to the experimental vessel itself, they develop regardless of the intended selective pressures in any given experiment. The evolution of wall growth and cell aggregation inside the continuous culture vessel seeds competing subpopulations, differentially restricting nutrient access for aggregating cells and skewing the likelihood of dilution (Smukalla et al. 2008; Fekih-Salem et al. 2013), both variables that should be fixed. Thus, developing a strain with reduced potential for evolving biofilm-related traits in this type of experimental system has many practical benefits.

Combating biofilm-related traits is also important in industry and medicine. Biofilm formation is a cell-surface adhesion trait that enables pathogenic organisms to persist on the surfaces of medical devices and even colonize human tissues (Kojic and Darouiche 2004; Verstrepen, Reynolds, and Fink 2004). Flocculation, a related cell-cell adhesion phenotype (Guo et al. 2000), is a mechanism by which yeast can survive stresses including treatment from antimicrobial compounds (Smukalla et al. 2008; Stratford 1992), with the cells on the interior of the floc physically protected from chemical treatments that more easily kill the outer layer of cells. This makes flocculation a problematic trait from a health perspective, and illustrates the importance of better characterizing the genetic basis of complex biofilm-related traits.

Cell aggregation, which we define here as an umbrella term to include both flocculation and mother/daughter separation defects (Stratford 1992), has dozens of known contributing genes identified by QTL mapping, deletion collection, genetic screen, and linkage analysis studies (Lee, Magwene, and Brem 2011; Granek et al. 2013; Brem et al. 2002; Ryan et al. 2012; Taylor and Ehrenreich 2014; H. Y. Kim et al. 2014; Roop and Brem 2013; Borneman et al. 2006; Ratcliff et al. 2015; Palecek, Parikh, and Kron 2000; Verstrepen et al. 2005; Cullen 2015; Taylor et al. 2016; Brückner and Mösch 2012). Given the extensive list of genes involved in aggregation that could potentially contribute to its evolution in a chemostat, our primary interests in this study were determining across many evolution experiments whether the genes involved in the evolution of aggregation were expected or novel, and ascertaining whether aggregating clones all achieved this final phenotype through one primary or many equally favored adaptive routes.

To ask how yeast evolve aggregation, we used multiplexed parallel evolution experiments coupled with genetics and whole genome sequencing to determine the causal mutations in 23 aggregating clones isolated from evolution experiments that ran 300 or more generations. Despite the known genetic complexity of aggregation, most of the causal mutations appeared to operate through a favored adaptive route: activating flocculation gene *FLO1*. Blocking this favored route by deleting *FLO1* significantly reduced incidence of flocculation in further evolution experiments, demonstrating the efficacy and potential of data-driven strain engineering, even for complex traits.

3.2 Results

Experimental evolution studies using continuous culture systems have suffered from small sample sizes in the past, a challenge that has been addressed through our multiplexed miniature chemostat system (Miller et al. 2013). In a previous study designed to test changes in fitness in

response to different nutrient limitations, we ran 96 miniature chemostats under three different nutrient limitations for 300 generations (Miller, A.W., in preparation). We observed that by 300 generations 34.7% had gained a visible cell aggregation phenotype not present in the ancestral strain, an S288C derivative that cannot aggregate due to a nonsense mutation in transcription factor Flo8 (Liu, Styles, and Fink 1996).

3.2.1 *Majority of aggregating clones demonstrate characteristics of true flocculation*

We selected clones for further study from the 23 populations with a strong aggregation phenotype. We conducted a number of phenotypic and genotypic analyses on the selected clones in order to determine how each strain had independently evolved the ability to aggregate. We quantified the settling ability of the evolved clones compared to the ancestral strain (Figure 3.1), a metric that describes the primary phenotype of interest in these experiments.

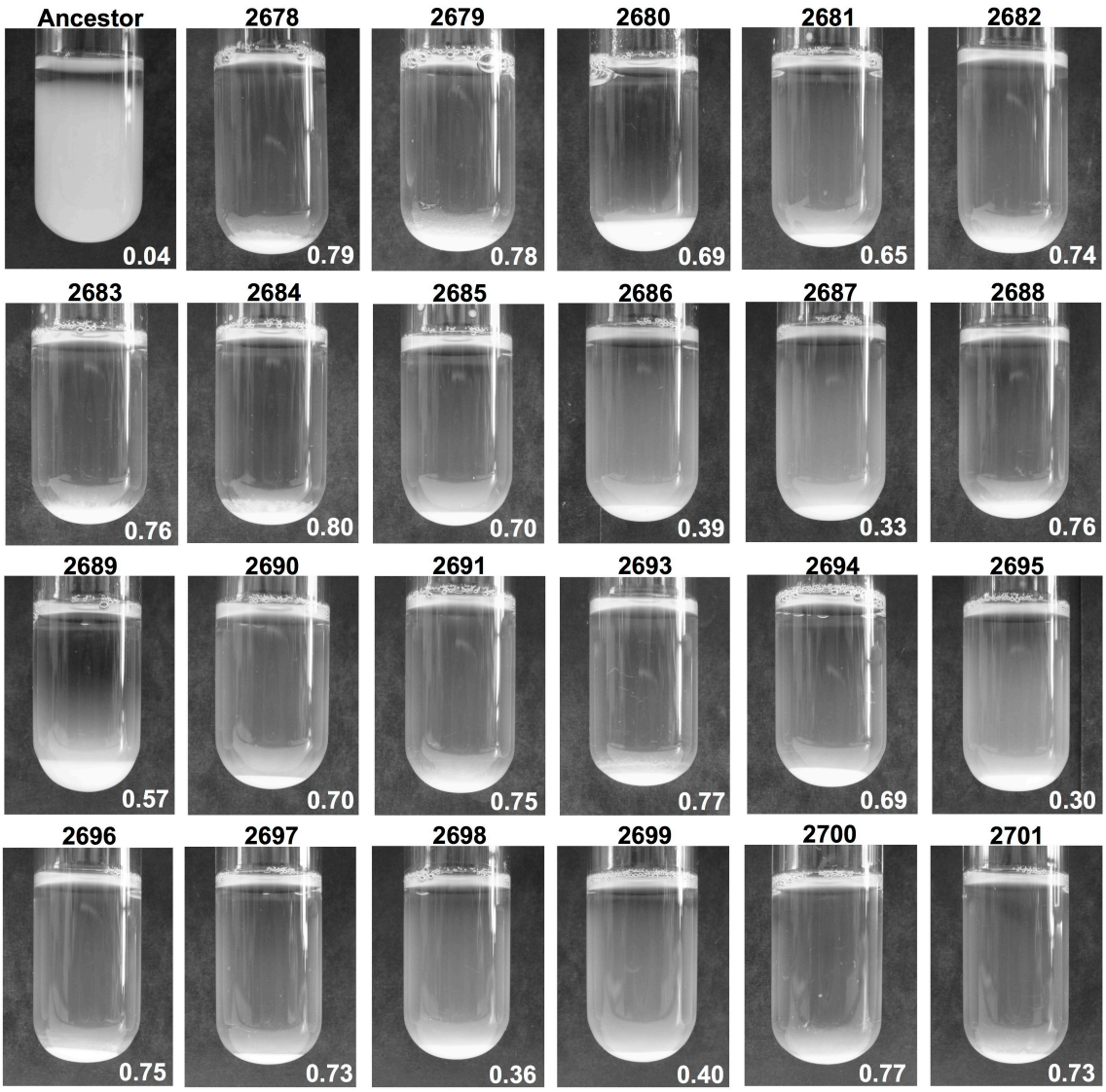


Figure 3.1. Quantitative settling of aggregating evolved clones. Images of the 60-minute settling time point for ancestral strain FY4 and 23 evolved clones with aggregation trait. Cultures were grown to saturation in 5mL YEPD liquid media. Settling ratio values are shown in bottom right of each image; ratios are the mean of three measurement replicates on the image shown.

We also examined the cellular morphology of all evolved clones microscopically and determined that two of the 23 clones (YMD2680 and YMD2689) show a cellular chaining phenotype indicative of a mother-daughter separation defect, while the remaining clones had aggregating round cells characteristic of cell-cell adhesion and true flocculation (Figure 3.2).

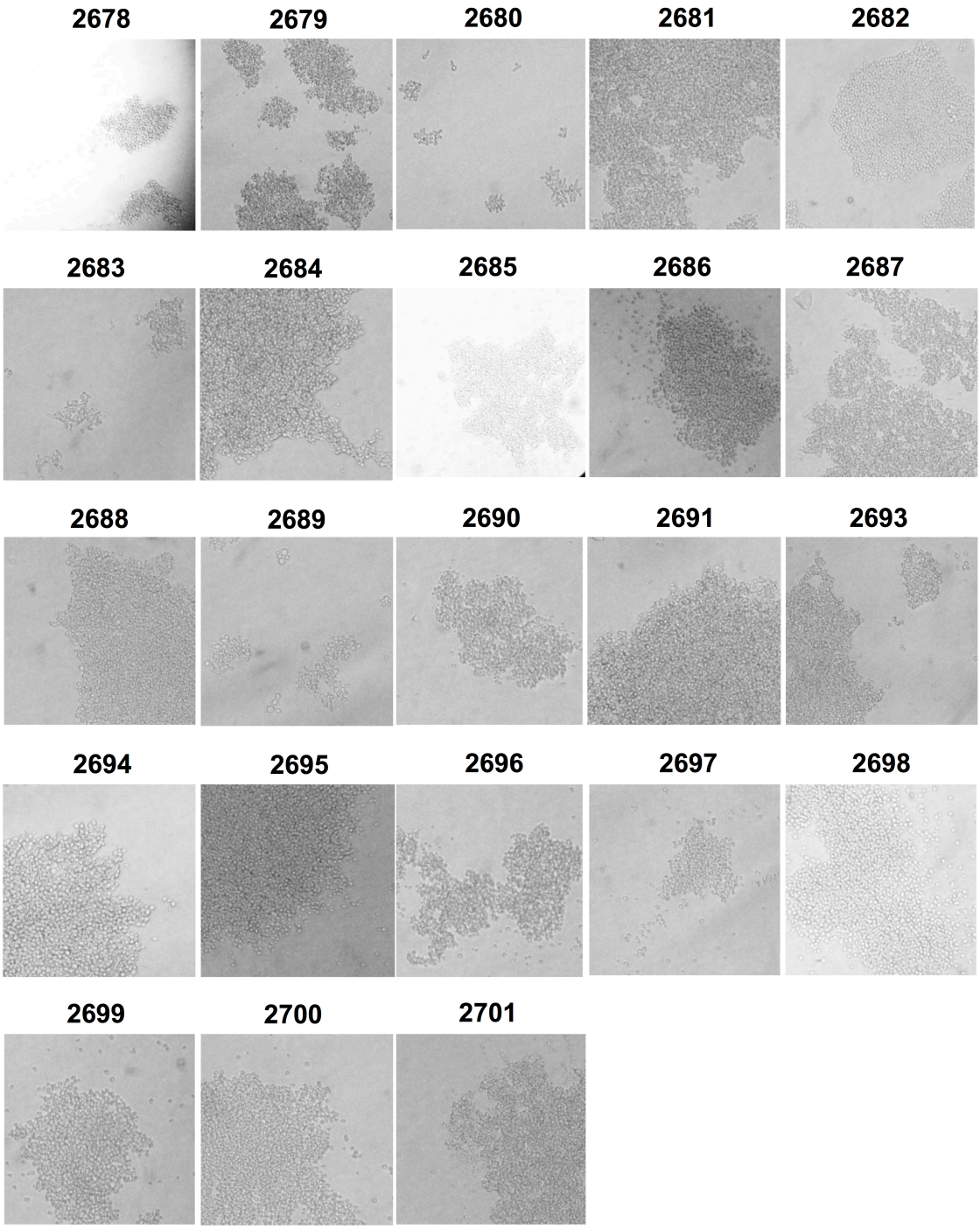


Figure 3.2. Micrographs at 150X of 23 evolved clones of *S. cerevisiae* with aggregation phenotype. Micrographs received additional processing (grey scale conversion, 20% increase in brightness, 20% increase in contrast) to better highlight the phenotypes. YMD2680 and YMD2689 exhibit mother-daughter separation defects while all other clones exhibit flocculation.

We confirmed the bud separation defect in YMD2680 and 2689 using calcofluor white staining (Figure 3.3), which preferentially stains the increased chitin present at yeast bud scars (Pringle 1991).

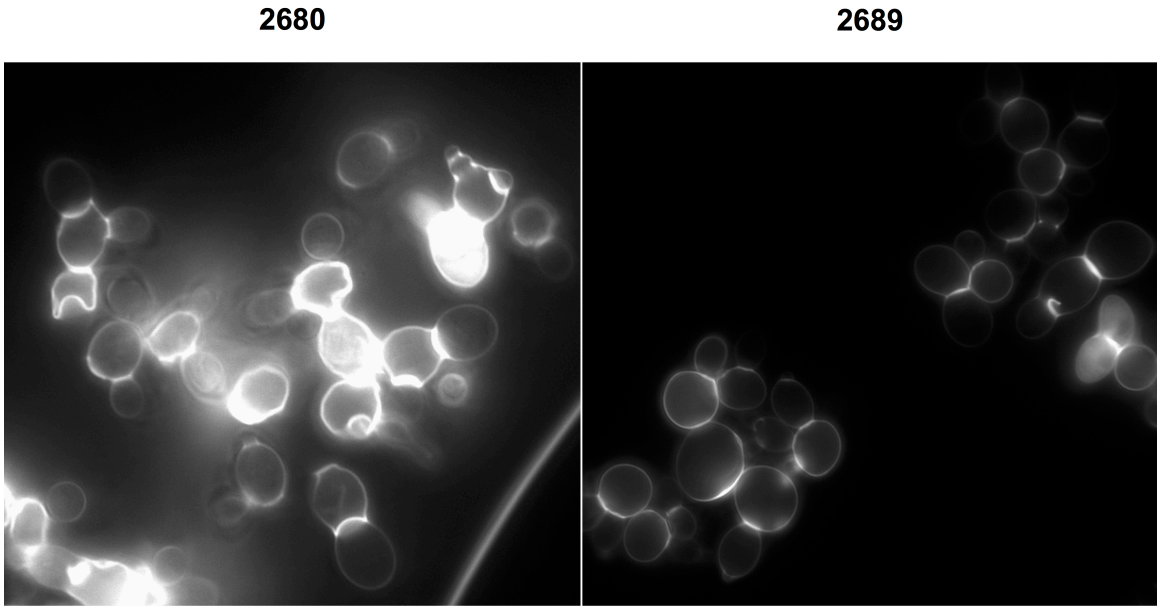
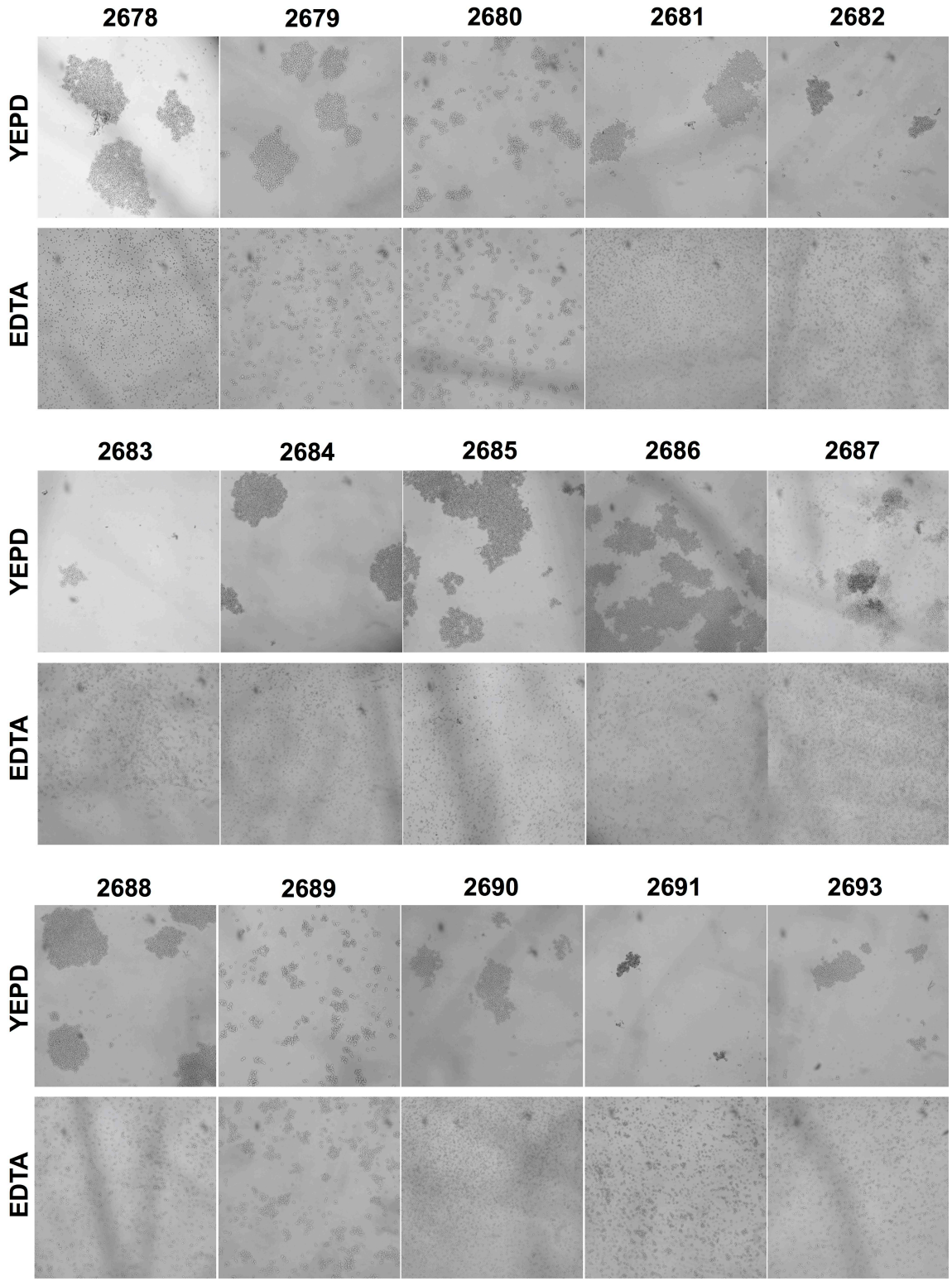


Figure 3.3. DAPI filter micrographs of evolved clones YMD2680 and 2689 with calcofluor white staining. Staining at bud scars highlights the mother-daughter separation defect in these strains at 630X magnification.

To further distinguish separation defects from flocculation, we treated the evolved clones with a de-flocculation buffer containing a chelating agent, EDTA; true flocculation is facilitated by calcium ions and reversible, while separation defects are not (Stratford 1989; Liu, Styles, and Fink 1996). We verified that all clones excluding YMD2680 and 2689 exhibit true flocculation that is reversible upon treatment with EDTA (Figure 3.4).


2678
2679
2680
2681
2682
YEPD
EDTA
2683
2684
2685
2686
2687
YEPD
EDTA
2688
2689
2690
2691
2693
YEPD
EDTA

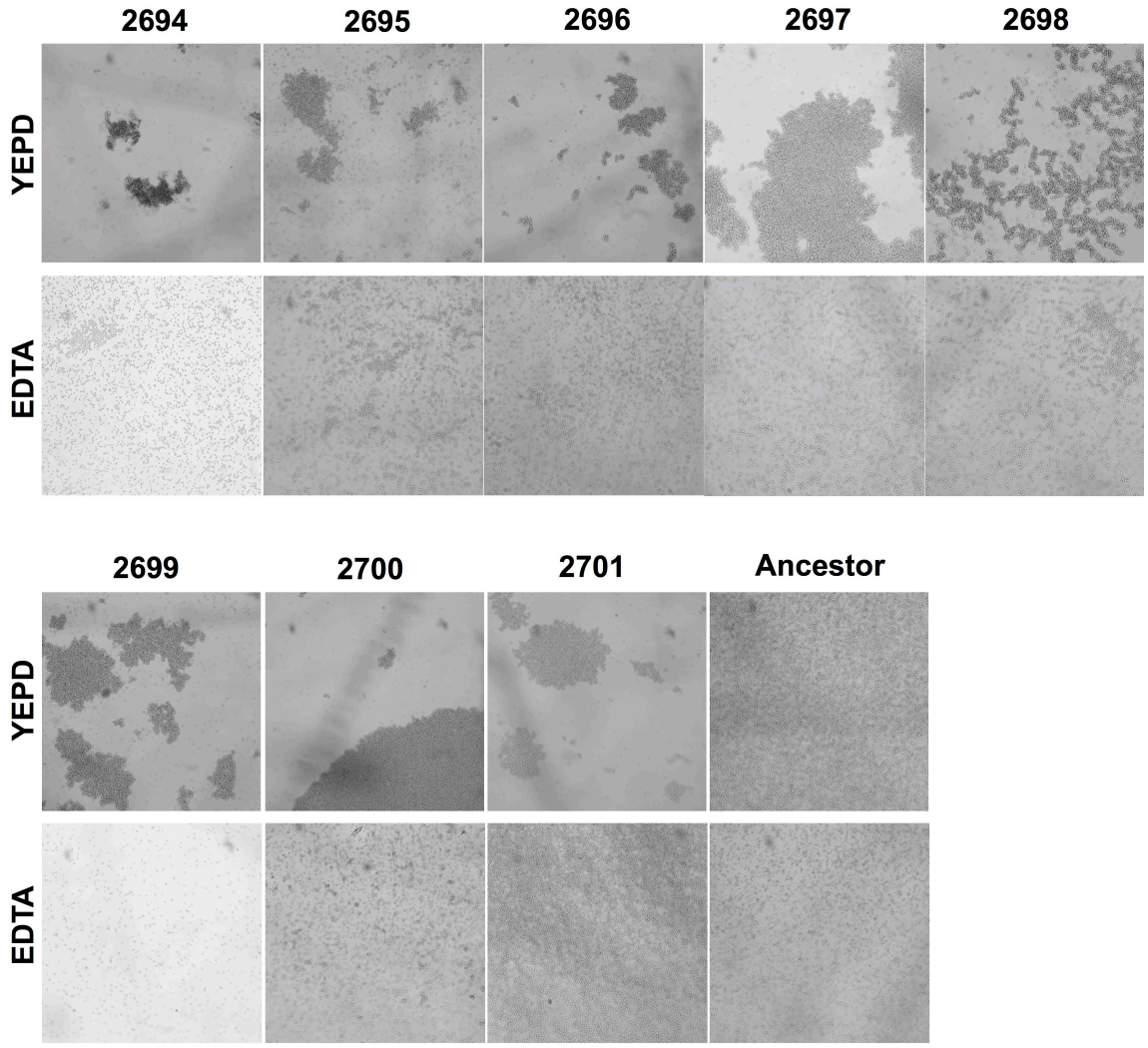


Figure 3.4. EDTA treatment of evolved clones stratifies mother-daughter separation from true flocculation. Micrographs received additional processing (grey scale conversion, 20% increase in brightness, 20% increase in contrast) to better highlight the phenotypes. Aggregates caused by separation defects fail to disperse following treatment with 4mM EDTA.

3.2.2 *Mutations in* FLO1 *promoter and genes* TUP1 *and* ACE2 *are primary adaptive routes to aggregation*

We performed Whole Genome Sequencing (WGS) on the 23 clones from generation 300 of the evolution experiments and analyzed the resulting sequence data to identify Single Nucleotide Variants (SNVs), small insertions or deletions (indels), Copy Number Variants (CNVs), and structural variants (Table S2, Materials and Methods). We developed a list of candidate genes

likely to contribute to the evolution of aggregation phenotypes (Table 3.1) from 17 different papers examining biofilm and cell aggregation related-traits, and several of the SNVs identified in our clones were in candidate genes (e.g. *ACE2, HOG1, TUP1).* We did not identify any instances of reversion of the ancestral point mutation in transcription factor gene *FLO8.*

Table 3.1 Aggregation candidate genes.

GENE	SYSTEMATIC NAME	STUDY
ACE2	*YLR131C*	Ratcliff 2015
AGA1	*YNR044W*	Brückner and Mösch 2012
AMN1	*YBR158W*	Li 2013
ASH1	*YKL185W*	Brückner and Mösch 2012
CDC28	*YBR160W*	Lee 2011
CYC8	*YBR112C*	Brückner and Mösch 2012
CYR1	*YJL005W*	Granek 2013
DEP1	*YAL013W*	Brückner and Mösch 2012
DIA1	*YMR316W*	Palecek 2000
END3	*YNL084C*	Taylor 2014
FIG2	*YCR089W*	Brückner and Mösch 2012
FLO1	*YAR050W*	Brem 2002
FLO10	*YKR102W*	Brückner and Mösch 2012
FLO11	*YIR019C*	Granek 2013, Ryan 2012
FLO1p	*YAR062W*	Vestrepen 2005
FLO5	*YHR211W*	Brückner and Mösch 2012
FLO8	*YER109C*	Granek 2013, Brem 2002, Taylor 2014, Ryan 2012
FLO9	*YAL063C*	Roop 2013
FUS3	*YBL016W*	Brückner and Mösch 2012
GCN1	*YGL195W*	Granek 2013
GCN2	*YDR283C*	Brückner and Mösch 2012
GCN4	*YEL009C*	Brückner and Mösch 2012
GLN3	*YER040W*	Brückner and Mösch 2012
GPB1	*YOR371C*	Taylor 2016
HAA1	*YPR008W*	Brückner and Mösch 2012
HDA1	*YNL021W*	Brückner and Mösch 2012
HOG1	*YLR113W*	Cullen 2015
HOT1	*YMR172W*	Granek 2013
IRA1	*YBR140C*	Roop 2013, Taylor 2016
IRA2	*YOL081W*	Roop 2013, Taylor 2014, Taylor 2016
IRC8	*YJL051W*	Taylor 2016
KSS1	*YGR040W*	Brückner and Mösch 2012

MFG1	*YDL233W*	Ryan 2012
MGA1	*YGR249W*	Borneman 2006, Brückner and Mösch 2012
MSB2	*YGR014W*	Brückner and Mösch 2012
MSN2	*YMR037C*	Granek 2013
MSS11	*YMR164C*	Su 2009, Kim 2004, Kim 2014, Taylor 2014, Ryan 2012
NRG1	*YDR043C*	Brückner and Mösch 2012
NRG2	*YBR066C*	Brückner and Mösch 2012
PGU1	*YJR153W*	Cullen 2015
PHD1	*YKL043W*	Borneman 2006, Brückner and Mösch 2012
PPM1	*YDR435C*	Granek 2013
PRP42	*YDR235W*	Granek 2013
RGA1	*YOR127W*	Li 2013
RGT1	*YKL038W*	Granek 2013
RIM101	*YHL027W*	Brückner and Mösch 2012
RME1	*YGR044C*	Brückner and Mösch 2012
SFL1	*YOR140W*	Brückner and Mösch 2012, Taylor 2016
SKS1	*YPL026C*	Granek 2013
SLF1	*YDR515W*	Granek 2013
SNF1	*YDR477W*	Bruckner and Mosch 2012
SOK2	*YMR016C*	Borneman 2006, Brückner and Mösch 2012
SOL3	*YHR163W*	Granek 2013
SSN3	*YPL042C*	Taylor 2016
SSN8	*YNL025C*	Taylor 2016
STA1	*STA1*	Kim 2004, Kim 2014
STE12	*YHR084W*	Kim 2004, Brückner and Mösch 2012
SWI5	*YDR146C*	Brückner and Mösch 2012
TEC1	*YBR083W*	Kim 2004, Brückner and Mösch 2012
TPK1	*YJL164C*	Brückner and Mösch 2012
TPK2	*YPL203W*	Brückner and Mösch 2012
TRR1	*YDR353W*	Taylor 2014
TUP1	*YCR084C*	Brückner and Mösch 2012
URE2	*YNL229C*	Brückner and Mösch 2012
YAK1	*YJL141C*	Granek 2013, Brückner and Mösch 2012
YAP1	*YML007W*	Brückner and Mösch 2012

Curated list of candidate genes with known contributions to separation defects, flocculation, or other biofilm-related phenotypes and their publication(s) of origin.

In both clones harboring separation defects, we discovered short insertions and deletions in the transcription factor gene *ACE2*, both of which cause a shift in the reading frame and introduction of a premature stop codon. These results are consistent with prior literature showing that loss of

function mutations in this gene cause settling/clumping phenotypes in other experimental evolution scenarios (Ratcliff et al. 2015; Voth et al. 2005; Oud et al. 2013; Koschwanez, Foster, and Murray 2013). Furthermore, *ace2* null mutants have the characteristic cell separation defect that we observed in our clones. We consider this *ACE2* mutation causative of the aggregation phenotype in these two clones (Table 3.2), with possible modification by *BEM2*, a gene involved in bud emergence that is also mutated in both clones (Bender and Pringle 1991; Kim et al. 1994).

In the 21 flocculent clones, the most common mutation we identified was a full-length insertion of a yeast transposable (Ty) element in the promoter region of *FLO1*. We saw this insertion in 12 of our clones, distributed in two hotspot regions between 95 and 156 bp and 394 and 470 bp upstream of the *FLO1* start codon (Figure 3.5, with regulatory information from (Fichtner, Schulze, and Braus 2007; Fleming et al. 2014; Basehoar, Zanton, and Pugh 2004)). Sequence analysis narrowed the type of Ty element in these insertions to Ty1 or Ty2, and diagnostic PCR and restriction digestion of nine inserts confirmed they were all Ty1 elements. In *FLO1* overexpression, localization, and deletion studies, *FLO1* has been shown to cause flocculation (Guo et al. 2000; Bony, Barre, and Blondin 1998; Smukalla et al. 2008); notably, Smukalla et al demonstrated that GAL-induced expression of *FLO1* in S288C, the background strain for these evolution experiments, induces flocculation, which supports the role we observe for *FLO1* regulation.

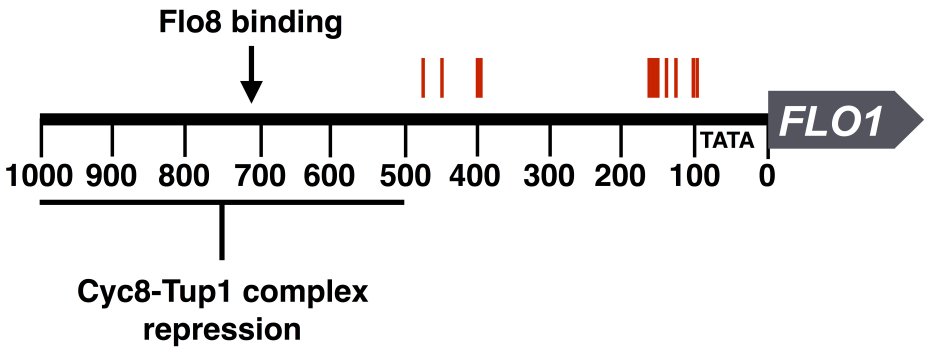


Figure 3.5. Ty element insertion sites cluster in two regions of the *FLO1* promoter. The region 1kb upstream of *FLO1* is shown with the insertion site positions of Ty elements observed in 12 evolved clones in red. Locations shown in this figure serve to demonstrate the primary regions of insertion only; for best estimates of exact insertion locations see Table 1. Flo8 binding site and Tup1-Cyc8 repression information adapted from (Fichtner, Schulze, and Braus 2007; Fleming et al. 2014). The TATA box is shown at 96 bp from the start of the open reading frame, as in (Basehoar, Zanton, and Pugh 2004).

In the remaining nine clones, we identified several SNVs and larger insertions and deletions in candidate genes, including *TUP1*, *FLO9, IRA1,* and *HOG1*, and many more in non-candidate genes. Five clones harbored likely loss-of-function mutations in candidate *TUP1*: two stop-gained SNVs in clones YMD2679 and YMD2693; one 27 bp deletion in YMD2700; one 100 bp deletion in YMD2682; and one Ty element insertion in YMD2688. *TUP1* is a general repressor (Carrico and Zitomer 1998; Z. Zhang, Varanasi, and Trumbly 2002), but also a repressor of *FLO1* (Fleming et al. 2014), and loss-of-function mutations in this gene have been associated with flocculation since 1980 (Williams and Trumbly 1990; Teunissen, van den Berg, and Steensma 1995; Lipke and Hull-Pillsbury 1984; Stark, Fugit, and Mowshowitz 1980) potentially driven by interactions with the co-repressing RNA Polymerase II mediator complex (Fleming et al. 2014). The frequently observed mutations in *TUP1* could function to de-repress *FLO1* or any number of other candidate genes. In a different clone, YMD2695, we identified a 6.2 kb deletion

from 229 bp to 6.4 kb upstream of flocculin gene *FLO9*. We also identified high confidence mutations not previously associated with aggregation in nearly all clones.

3.2.3 *Bulk Segregant Analysis verifies causal mutations in novel genes*

Because of the number of high confidence mutations in each clone, we could make hypotheses about causality. To test causality and examine the genetic complexity of the trait in each clone, we turned to a different method, bulk segregant analysis (BSA). We backcrossed the 21 evolved flocculent clones to a non-flocculent strain isogenic to the ancestor but of the opposite mating type. We excluded the two clones with separation defects because their causality was clear and their budding defect interfered with tetrad dissection. BSA leverages meiotic recombination and independent assortment to link a trait to a causal allele, which will be observed in all progeny with the phenotype of interest. In turn, unlinked non-causal alleles should assort equally between progeny with and without the phenotype (Brauer et al. 2006; Birkeland et al. 2010; Segrè, Murray, and Leu 2006) (Figure 3.6A). Backcrossing also allowed us to estimate the genetic complexity of the trait: if two of four meiotic progeny have the phenotype and two do not, this indicates a single causal allele for the phenotype. We observed this 2:2 segregation pattern in 20 of the 21 evolved clones, and pooled and sequenced the progeny with and without the trait to identify which of the initial candidate alleles was causal.

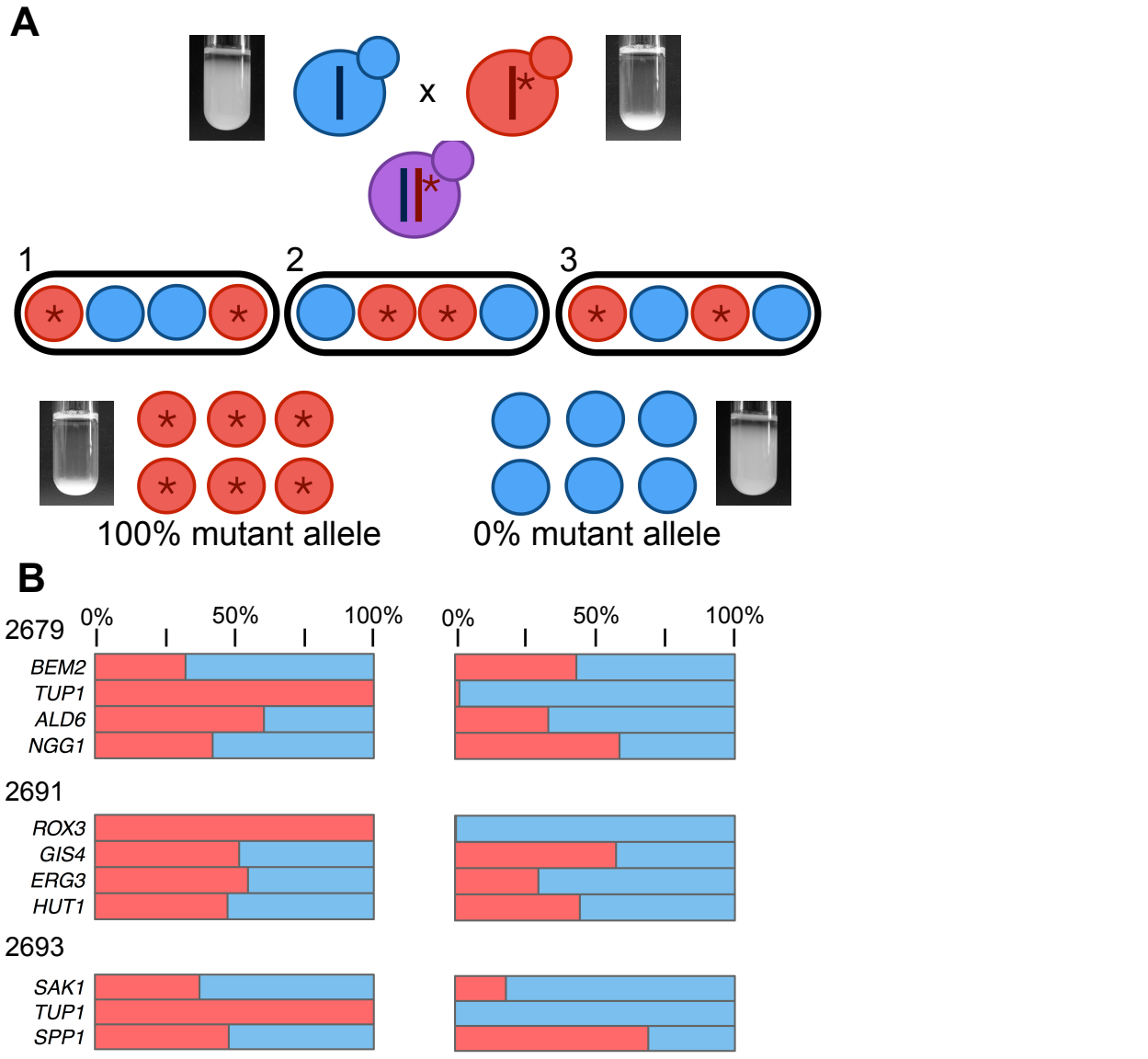


Figure 3.6. Bulk segregant analysis leverages recombination to identify mutations that co-segregate with the flocculation trait. A) An evolved clone with the phenotype of interest, shown here as settling in liquid culture, is backcrossed to the ancestral strain lacking the phenotype. Dissection of tetrads resulting from this cross reveals the segregation pattern of the trait among meiotic progeny, with 2:2 segregation (two segregants with the settling trait and two without) indicative of single gene control of the trait. Segregants with and without the trait are pooled and sequenced, and alleles that co-segregate with the trait are identified as causal. **B)** For three backcrosses, pooled sequencing results are shown for both pools of segregants, those with the settling trait on the left, and without the settling trait on the right. The strain identifier for the evolved clone in the cross is shown on the left, along with a list of candidate genes that had high quality Single Nucleotide Variant calls in the clone. The red bar shows the % of each of those candidate mutations seen in each pool, with mutations seen at 100% frequency identified as co-segregating with the trait and therefore causal.

We subjected all of our clones with genic mutations, including large insertions and deletions, to BSA analysis, and included four of the clones with a Ty element insertion in the *FLO1* promoter. The analysis pipeline (Materials and Methods) identified mutations at 100% frequency in the flocculent pools, and confirmed the causality of the *FLO1, TUP1, FLO9,* and *ROX3* mutations. BSA also confirmed the causality of mutations in *CSE2* and *MIT1,* genes not previously associated with flocculation (though both have been linked to related traits such as invasive growth and biofilm formation, see below). For the three evolved clones with causal SNVs, the frequency of each candidate in the flocculent and non-flocculent pools is shown in Figure 3.6B and 3.6C; in each case, the causal mutant allele was at 100% frequency in the flocculent pool. Using the combined results of WGS and BSA, we were able to resolve the causal mutation for all 23 of the evolved clones, with a complete summary of our findings in Table 3.2.

Table 3.2 Causal mutations for the aggregation phenotype in 23 evolved clones.

CLONE (YMD)	NUTRIENT LIMITATION	CAUSAL GENE	SYSTEMATIC NAME	MUTATION TYPE
2678	G	*CSE2*	*YNR010W*	Ty insertion in ORF at R137
2679	S	*TUP1*	*YCR084C*	Stop-gained Q181*
2680	G	*ACE2*	*YLR131C*	S115 indel (2bp deletion); premature stop introduced
2681	G	*FLO1*	*YAR050W*	Ty in promoter (156bp upstream of ORF)
2682	P	*TUP1*	*YCR084C*	Q107-P143 deletion in ORF (106bp); premature stop introduced
2683	G	*FLO1*	*YAR050W*	Ty in promoter (139bp)
2684	S	*FLO1*	*YAR050W*	Ty in promoter (127bp)
2685	G	*FLO1*	*YAR050W*	Ty in promoter (397bp)
2686	P	*FLO1*	*YAR050W*	Ty in promoter (151bp)
2687	P	*FLO1*	*YAR050W*	Ty in promoter (156bp)
2688	P	*TUP1*	*YCR084C*	Ty insertion in ORF at L341
2689	G	*ACE2*	*YLR131C*	L192indel (1bp insertion); premature stop introduced
2690	G	*FLO1*	*YAR050W*	Ty in promoter (449bp)
2691	G	*ROX3*	*YBL093C*	Stop-gained C138*
2693	G	*TUP1*	*YCR084C*	Stop-gained Q101*
2694	S	*MIT1*	*YEL007W*	L552-M585 deletion in ORF (101bp); premature stop introduced
2695	P	*FLO9*	*YAL063C*	6.2kb deletion in promoter (229bp upstream of ORF)
2696	G	*FLO1*	*YAR050W*	Ty in promoter (470bp)
2697	P	*FLO1*	*YAR050W*	Ty in promoter (95bp)
2698	G	*FLO1*	*YAR050W*	Ty in promoter (394bp)
2699	P	*FLO1*	*YAR050W*	Ty in promoter (102bp)
2700	P	*TUP1*	*YCR084C*	V592 indel (27bp deletion); in frame
2701	P	*FLO1*	*YAR050W*	Ty in promoter (152bp)

The gene in which or in front of which the causal mutation was found is identified here, along with the type of mutation we recorded. Also shown is the nutrient limitation in which the clones were evolved: G, S, or P for glucose-limited, sulfate-limited, and phosphate-limited, respectively. Positional information about SNVs and indels is exact; other values shown are approximate (Materials and Methods).

3.2.4 *Functional* FLO1 *is necessary for flocculation driven by* ROX3*,* CSE2*, and* MIT1 *mutations*

Given the large number of potentially activating mutations that we recovered in *FLO1*, we hypothesized that the causal *ROX3* and *CSE2* mutations we recorded also act through *FLO1*, via loss of repression. Several lines of evidence make *ROX3* a reasonable candidate repressor for

FLO1 and/or other *FLO* genes. Loss-of-function mutations in *ROX3* have been previously associated with flocculation (Brown, Evangelista, and Trumpower 1995) and also pseudohyphal growth, which is a trait related to haploid invasion and regulated by *FLO* genes (Guo et al. 2000). *ROX3* and *CSE2* both encode components of the RNA polymerase II mediator complex, which also includes Sin4, Srb8, and Ssn8, whose role in *FLO* gene repression is described in Fichtner *et al* (Fichtner, Schulze, and Braus 2007). Mutations in other components of Mediator have previously been shown to cause clumping (Koschwanez, Foster, and Murray 2013) and there are known interactions between Mediator and repressor *TUP1* (Papamichos-Cronakis et al. 2000; Fleming et al. 2014), in which we also recovered mutants. In order to test the relationship between the *ROX3* and *CSE2* mutations and *FLO1,* we examined the ratio of settling to non-settling progeny in crosses between a *flo1* knockout strain and strains harboring the *CSE2* and *ROX3* causal mutations. 50% settling and 50% non-settling segregants compiled over all tetrads would indicate, for example, that both the *cse2 FLO1* and *cse2 flo1* segregants flocculate and that the function of the *cse2* mutation is not dependent on a functional *FLO1*. 25% settling and 75% non-settling segregants, and the presence of tetrads segregating 1:3 and 0:4, would indicate that the double mutant does not flocculate and a functional *FLO1* is required for the effect of the *cse2* (or *rox3*) mutation to be observed. We observed that the double mutants show a wild-type, non-flocculent settling phenotype, i.e., that *flo1* is epistatic to the other mutations. This indicates that *FLO1* is required for these mutations to have an effect and lends support to the hypothesis that Rox3 and Cse2 function as *FLO1* repressors in the wild-type strain.

Analysis of progeny with a *flo1* null mutation and the *MIT1* allele from YMD2694 revealed similarly that the *MIT1* mutation requires a functional *FLO1* to cause flocculation. *MIT1* is a known transcriptional regulator of flocculin genes *FLO1*, *FLO10*, and *FLO11*, and null

mutants of *MIT1* exhibit reductions in hallmark biofilm-related traits including invasive and psuedohyphal growth and colony complexity (Cain et al. 2012), which are related to flocculation in S288C (Liu, Styles, and Fink 1996; Fichtner, Schulze, and Braus 2007). This role of *MIT1* in the literature suggests that the deletion we record is not a loss-of-function mutation, although it is not dominant (data not shown). If the *MIT1* deletion in YMD2694 caused loss of function, we would expect to see a non-flocculent phenotype, as we confirmed is observed in a *mit1* deletion strain; instead, the deletion causes a flocculation phenotype, indicating that it serves in some way to enhance the function of *MIT1*. The deletion itself is out of frame and therefore results in a modified C-terminus of the protein, including a premature stop codon and truncation of the final product. From the extensive literature on the *MIT1* ortholog in *Candida albicans*, *WOR1*, we know that DNA binding activity is likely confined to the N-terminal portion of the protein, far from the mutation in this allele of *MIT1:* in *WOR1,* two DNA binding regions in the N-terminal portion of the protein are sufficient for full activity (Lohse et al. 2010; S. Zhang et al. 2014). *WOR1* and *MIT1* also both have a self-regulatory mechanism through a positive feedback loop, a potential mechanism for the enhanced function implicated by the mutation we observe (Cain et al. 2012; Zordan, Galgoczy, and Johnson 2006).

3.2.5 *Phenotypic variation suggests secondary modifiers influence flocculation*

Though we identified the *FLO1* promoter Ty element insertion as the primary causal allele for the aggregation trait in 12 of our clones, we observed variation in the types of flocs produced in our preliminary microscopy of the clones (Figure 3.2), and differences in settling even among all strains with a *FLO1* promoter insertion. These differences were not caused by Ty element direction, proximity, or type: all of the Ty elements we were able to validate with PCR and restriction site polymorphisms were of type Ty1. Secondary genetic modifiers of the flocculation

trait are an alternative explanation for this phenotypic variation. To identify strains potentially carrying secondary modifier mutations, we examined the distribution of quantitative settling ratios across a subset of settling segregants for each cross. Segregants without a modifier were expected to match the evolved parent settling phenotype, while a distribution of settling abilities would be seen as evidence of a potential modifier (Figure 3.7).

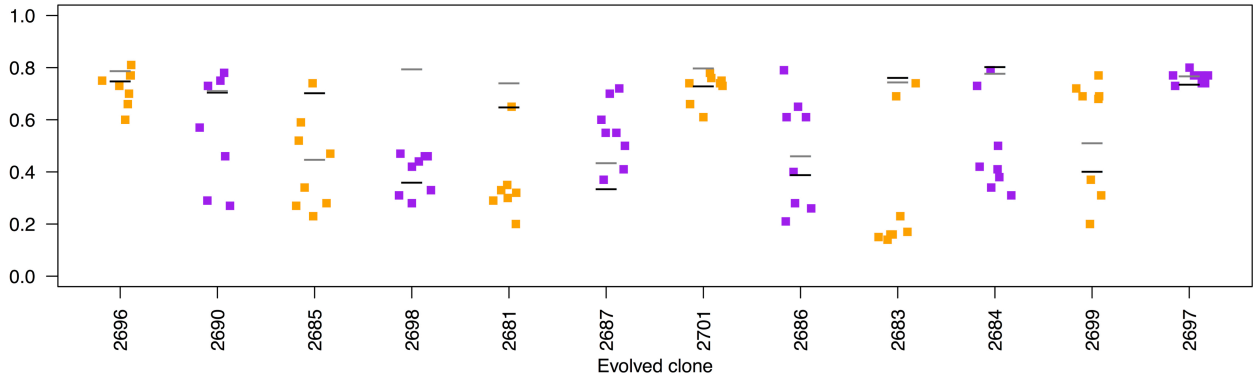


Figure 3.7. Settling ratios harbor quantitative variation among clones with *FLO1* promoter Ty insertions. Distribution of settling ratios from 0 to 1 among settling segregants from backcrosses using evolved clones with *FLO1* promoter Ty insertions. Variation among segregants indicates additional genetic modifiers of the phenotype beyond the single causal allele identified through BSA and WGS. Clone YMD2683 has a verified secondary modifier and a distribution of settling segregants into two groups. Horizontal lines represent independent biological replicates for the evolved clone contributing to each cross, each a mean of three measurement replicates.

One strong candidate for multiple alleles contributing to the aggregation phenotype was clone YMD2701, the only evolved clone that did not segregate the settling phenotype 2:2 during BSA. Sequencing analysis revealed this clone does have the *FLO1* promoter Ty1 insertion. We also identified an amplification of chromosome I in this clone, both copies of which have the promoter insertion, indicating that two copies of the causal allele are segregating in this backcross; this genotype is consistent with the segregation pattern we observed (Figure 3.8). Within the segregant settling ratios, however, we did not observe this aneuploidy to be a modifier of the trait (Figure 3.7).

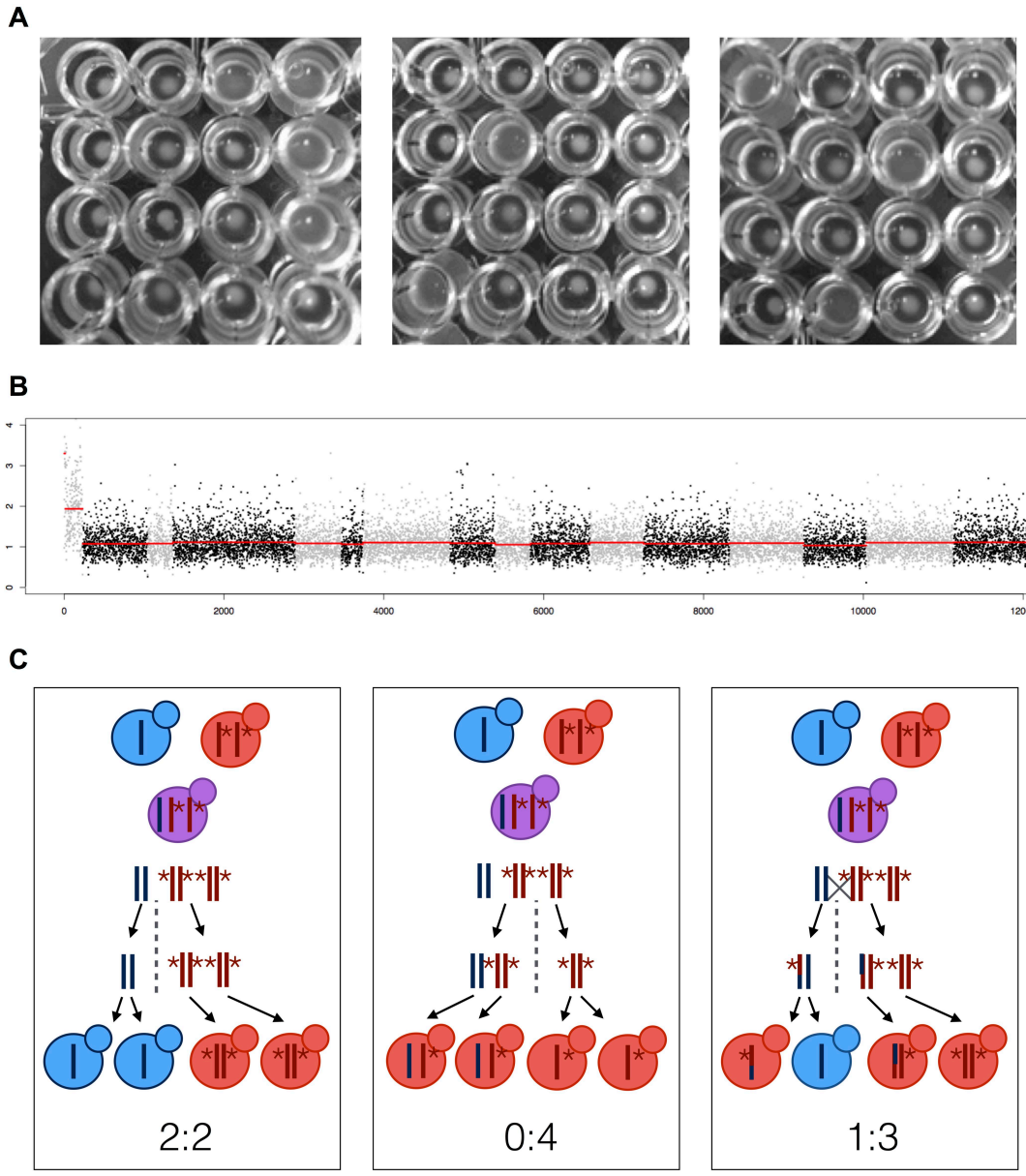
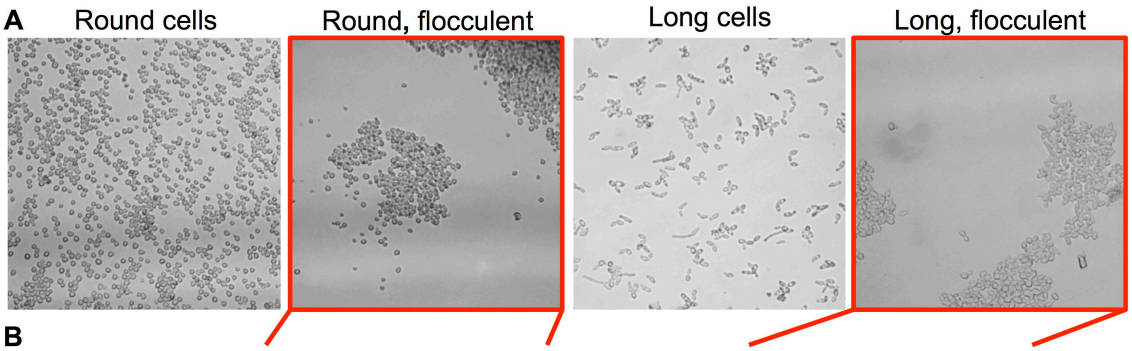


Figure 3.8. Segregation of settling trait in a backcross between evolved clone YMD2701 and lab strain FY5 is explained by chromosome I aneuploidy. A) In this cross, settling trait does not segregate 2:2. 16 representative tetrads are shown horizontally with segregation patterns of 2:2, 1:3, and 0:4 non-settling:settling. **B)** Copy number analysis of clone YMD2701 reveals a chromosome I amplification. Y-axis values represent relative copy number increase, from 1 copy lost (0) to 3 copies gained (4) from a haploid starting point. X-axis values are genome coordinate. Gene *FLO1* is on chromosome I, and clone YMD2701 is known to have a Ty insertion in the *FLO1* promoter. **C)** Three possible outcomes of meiosis with the causal *FLO1* Ty insertion on both copies of chromosome I leads to the observed segregation pattern.

In clone YMD2683, we identified a secondary modifier related to cell morphology. In our initial microscopy (Figure 3.2), we observed that clone YMD2683 had an unusual elongated cell morphology, which we observed segregating in the backcross as well. Microscopy of segregants from this cross revealed four phenotypic classes: round, suspended cells; round, flocculent cells; long, suspended cells; and long, flocculent cells (Figure 3.9A). Segregants from the backcross involving evolved clone YMD2683 had two different settling ratios, the weaker of which correlated with the round, flocculent morphology, while the stronger settling ratio correlated with the long, flocculent cell morphology (Figure 3.9A, C). WGS of the original clone identified high quality SNVs in genes *IRA1, HSL7, VTS1,* and *TCP1*. PCR and Sanger sequencing of each of these genes in segregants from each phenotypic class revealed co-segregating missense mutations in *HSL7* and *IRA1* in all segregants with the long cell phenotype (examples in Figure 3.9B), suggesting that one or both of these mutations is functioning as a secondary modifier to enhance the phenotype from the *FLO1* Ty element insertion. *HSL7* and *IRA1* are located only 13kb apart from each other on chromosome II, indicating that this co-segregation could be due to linkage rather than the contribution of both genes to the trait, though null mutations in each have been linked to abnormal cell morphology (*HSL7,* (Kucharczyk et al. 1999; Fujita et al. 1999)) and flocculation (*IRA1,* (Verstrepen, Reynolds, and Fink 2004; Halme et al. 2004)).



B

	Tetrad 1, spore 2	Tetrad 1, spore 4	Tetrad 2, spore 3	Tetrad 3, spore 1
FLO1 Ty	EV	EV	EV	EV
HSL7	WT	WT	EV	EV
IRA1	WT	WT	EV	EV
VTS1	WT	WT	WT	WT
TCP1	WT	EV	WT	EV

C

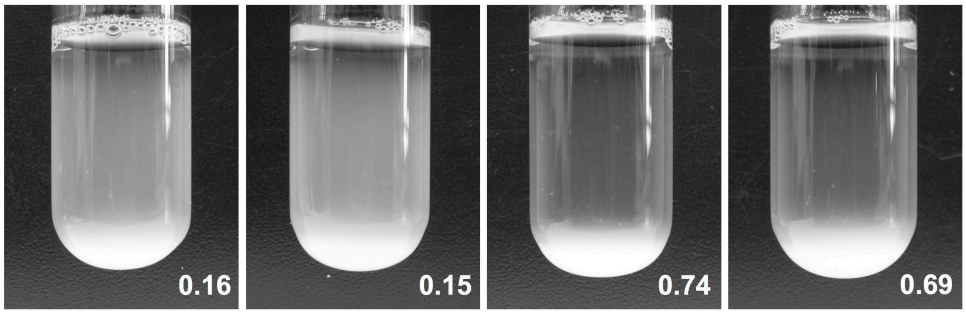


Figure 3.9. Evolved clone exhibits morphology-related secondary modifier of flocculation phenotype.

A) Meiotic segregants show four phenotypic classes combining morphology and flocculation. Micrographs received additional processing (grey scale conversion, 20% increase in brightness, 20% increase in contrast) to better highlight the phenotypes. **B)** Sequencing results for four example flocculent segregants are shown, from two of the phenotypic classes in 3.9A: two segregants have round flocculent cells, and two with long flocculent cells. All segregants have the *FLO1* Ty insertion. The four candidate SNVs from the evolved clone were Sanger sequenced in the flocculent progeny and the match to the WT (S288C reference) or mutant/evolved base (EV) is recorded. **C)** Settling images and ratios for the four flocculent segregants that provided the sequencing data in 3.9B.

Another promising candidate for a secondary modifier was clone YMD2690: segregants from the backcross with this clone showed considerable variation in settling ratios, and the clone harbored a premature stop codon in candidate gene *HOG1* (Table 3.1), although a Ty element in the *FLO1*

promoter was identified as the primary causal mutation. Using Sanger sequencing of settling segregants we determined that *HOG1* was not a secondary modifier of the trait. We conducted additional testing using primers from Zara *et al* (Zara et al. 2009) (Table 3.3) to target the repeat region in flocculin gene *FLO11* in clone YMD2690 and found evidence in this clone of a *FLO11* repeat expansion of approximately 1 kb in length. All flocculin genes have long arrays of internal tandem repeats (Verstrepen et al. 2005); expansions of the internal repeats in *FLO11* have been shown to cause phenotypic variability in biofilm-related traits, and natural isolates of yeast exhibit significant variation in the copy number of the repeats (Fidalgo, Barrales, and Jimenez 2008; Zara et al. 2009). However, this expansion also did not correlate with variations in strength of the segregant settling ratios, demonstrating that the presence of the *FLO11* expansion in addition to the Ty element insertion did not significantly affect the strength of the phenotype. PCR of all of the evolved clones revealed that only this clone had any evidence of repeat expansion in *FLO11*.

3.2.6 *Deleting* FLO1 *increases time to evolve flocculation and reveals alternate adaptive routes*

The results of our analyses of the evolved clones demonstrate a clear role for *FLO1* in the evolution of flocculation; not only do we see changes in the *FLO1* promoter, many of the other mutations we recorded are in genes encoding proteins that function to regulate *FLO1 (TUP1)* or participate in complexes that regulate *FLO1 (ROX3, CSE2)*. We hypothesized that changes in the regulation of *FLO1* cause the flocculation phenotype in nearly all of the evolved clones, and that deleting *FLO1* would be a promising route for slowing the evolution of flocculation. Deleting a combination of *FLO* genes has been previously employed as a method to try to make lab strains easier to work with in evolution studies (Voordeckers and Verstrepen 2015), and modification of

the *FLO1* promoter has been effectively employed in biological circuits controlling flocculation (Ellis, Wang, and Collins 2009); however, it is unknown if specifically deleting *FLO1* would be effective on its own. We constructed a *flo1* strain and evolved 32 chemostat vessels of wild-type concurrently with 32 chemostat vessels of the *flo1* knockout strain, in glucose limited media for over 250 generations. Two knockout and one wild-type vessel were lost to contamination after generation 200.

We monitored all vessels for evidence of aggregation and recorded eight wild-type and one knockout strain that developed aggregation during the course of the experiment, a statistically significant reduction ($p=0.01$, Fisher's Exact Test). In order to determine the mechanism of the single aggregating *flo1* population, we performed WGS of a clone and found a Ty element insertion in the promoter of *FLO9*. In addition to the single knockout clone, we sequenced four wild-type strains that also evolved aggregation in the course of the experiment. Two of these harbored *FLO1* promoter Ty element insertions; another had a stop-gained mutation in *NCP1* that has not been verified as causal; and another had a deletion in *MIT1* exactly matching the deletion identified in the clone from the previous series of evolution experiments.

3.2.7 FLO1 *deletion does not affect rate of evolution for unrelated traits*

We expected that deleting *FLO1* would not impact the rate of evolution for unrelated traits, including wall sticking and separation defects, two other traits we monitored during the knockout evolution experiments. Cell-surface adhesion traits are more often associated with expression of *FLO11* (Guo et al. 2000; Verstrepen and Klis 2006), and we would not expect the frequency of evolving separation defects to be affected by changes to flocculation genes. For 32 of the vessels across both genotypes we recorded the occurrence of some amount of wall sticking two days

before the final time point; eight of these we recorded as strong wall growth at the final evolution time point. The strong wall growth observations were split equally between WT and *flo1* knockout populations. Similarly, for mother-daughter separation defects, which we observed through microscopy of each of the final evolution time points, we recorded 12 strains with separation defects, five from the wild-type and six from the knockout evolution experiments, with an additional wild-type strain with an inconclusive microscopy phenotype (Appendix B.1).

To explore the genetic origins of the wall sticking trait, we isolated clones with a strong wall sticking phenotype from six different populations, four from knockout experiments and two from wild-type experiments. Under the microscope, we observed that all six wall sticking clones harbored a separation defect. To determine if these were all caused by loss of function alleles of *ACE2*, we performed a complementation test using the *ace2* strain from the yeast deletion collection (Giaever et al. 2002). We determined that a loss-of-function mutation in *ACE2* was responsible for both the wall sticking and mother-daughter separation defects in five of the six clones. Despite this relationship, we did not observe a strong connection between wall sticking and separation defects on the population level, with 10 strains having only a separation defect, five having only strong wall growth, and only four populations having both phenotypes. The mechanism by which loss of function in *ACE2* facilitates wall sticking remains undetermined.

3.3 Discussion

Previous studies have successfully leveraged experimental evolution to understand the genetic contributors to complex traits (Voordeckers and Verstrepen 2015; Leu and Murray 2006; C. J. Brown, Todd, and Rosenzweig 1998; Hong and Gresham 2014). Evolution experiments have also contributed significantly to our understanding of how genomes evolve and the types of mutations typically observed in yeast grown in chemostats, including SNVs, CNVs, aneuploidy,

and transposable element insertions (Dunham et al. 2002; Araya et al. 2010; Gresham et al. 2008; Adams, Julian 2004; Adams and Oeller 1986). In our study, we built on these concepts to identify the mutations contributing most to the evolution of cell aggregation, an industrially and medically relevant trait in addition to a practically useful one for facilitating laboratory work. We determined that in experimental evolution in continuous culture, loss-of-function mutations in *ACE2* are the most common contributors to the evolution of mother-daughter separation defects and wall growth, and mutations that change the regulation of *FLO1* are the most common evolutionary route to flocculation. The majority of causal mutations identified in this study occurred in candidate genes selected for involvement in aggregation traits based on previous literature, but two of the causal mutations were in genes not previously associated with flocculation (*CSE2*, *MIT1*). Both our identification of new genetic associations with flocculation and of one favored adaptive route to flocculation demonstrate the efficacy of using experimental evolution as a tool to better understand important complex traits.

This study also demonstrates the power of evolution experiments to determine which genes, among the many genes that are associated with complex traits like flocculation, most frequently contribute to adaptation under specific constraints. Despite the many possible candidates, we saw few of those genes identified in the evolved clones in this study. This finding is in keeping with other work in eukaryotes demonstrating favored adaptive responses, not just in the clear relationships between nutrient limitation and the amplification of nutrient transporters (C. J. Brown, Todd, and Rosenzweig 1998; Gresham et al. 2008) but also in response to stress treatments. In a more saturated screen we might start to see contributions from other candidate genes or pathways, but large screens have also revealed parallel adaptation. A study of 240 yeast strains under selective pressure from the antibiotic Nystatin revealed significant parallelism in

mutational response through a single pathway (A.C. Gerstein, Lo, and Otto 2012) similar to the parallelism we discovered through *FLO1* regulation. The contributions of other candidate genes might also be revealed in a scenario in which an engineered strain has been modified to take away the primary adaptive routes we observed.

The primary mechanism of evolution we observed, a Ty element insertion in the *FLO1* promoter region, likely activates *FLO1* expression similarly to previous Ty element systems (Rothstein and Sherman 1980; Errede et al. 1984; Coney and Roeder 1988). The reverse orientation of the Ty element with respect to the open reading frame that we observed in all of our clones is the most common activating arrangement (Servant, Pennetier, and Lesage 2008; Boeke 2016), and the role of transposable elements in driving adaptive mutations has been well documented in yeast and other organisms (Chao et al. 1983; Tenaillon et al. 2016; Wilke and Adams 1992). Despite discovering one primary mechanism for evolving flocculation, we also show evidence for other genetic contributors modifying and enhancing the phenotype we observe. There is quantitative variation among settling segregants from crosses with our evolved clones (Figure 3.7) particularly among strains with the *FLO1* promoter Ty element insertion, and we confirmed one example of a secondary modifier of the settling trait in clone YMD2683, in which a change in cell morphology enhanced the trait from the *FLO1* promoter Ty element insertion. Across other clones with trait variation there is potential to discover additional modifiers, both in the form of known candidate genes, including other *FLO* genes with internal tandem repeats, and in genes that have not previously been associated with flocculation.

Each causal mutation in our clones represents a new possible avenue for engineering to reduce aggregation. These could be simple changes, such as fusing genes like *ACE2* and *TUP1* that frequently acquire loss-of-function mutations to essential genes, or increasing their copy

number or strain ploidy to increase the likelihood of "masking" deleterious recessive mutations (Otto and Goldstein 1992). They can also be iterative: deleting *FLO9* in the *flo1* background could even further reduce evolution of flocculation. Alternative strategies include reducing the mutation rate of these nondesirable mutations. The frequency at which we observe activating Ty elements driving flocculation also suggests future experiments aimed at reducing Ty element expression or mobility could be fruitful. Promising routes for reducing the Ty burden in evolution experiments include inhibiting Ty1 transposition (Xu and Boeke 1991) or utilizing different background strains. There is evidence that strain background contributes significantly to the likelihood of evolving flocculation in chemostat experiments. *Saccharomyces uvarum,* a budding yeast related to *S. cerevisiae* and often used in interspecific hybrid studies, has only Ty4 elements in its genome (Liti et al. 2005) and evolves flocculation more slowly than *S. cerevisiae* in chemostat experiments (Heil et al. 2016; Sanchez et al. 2016). Not only do different species of yeast have different Ty element burdens, natural isolates of *Saccharomyces cerevisiae* also provide strain-specific differences in Ty element burden (Bleykasten-Grosshans, Friedrich, and Schacherer 2013; Dunn et al. 2012) and a reservoir of variation in evolutionary potential which will be useful in future evolution experiments for studying flocculation and other complex traits.

Over the past six decades, experimental evolution in chemostats with yeast and bacteria has provided valuable insights into evolutionary dynamics and has proven to be a powerful tool for understanding complex traits. Now, with the advent of modern sequencing technology and common strain engineering methods, experimental evolution represents a promising direction for designing and testing strains with reduced (or increased) evolutionary potential. Evolution is gaining popularity as a tool for engineering: as just a few examples, in 2002, Yokobayashi *et al* used directed evolution to improve the function of a rationally designed circuit driving a

fluorescent reporter (Yokobayashi, Weiss, and Arnold 2002), and evolutionary engineering is commonly used to improve carbon source utilization of industrial strains (Garcia Sanchez et al. 2010; Shen et al. 2012; Zhou et al. 2012). Evolution poses a challenge to strain engineering as well: loss, change, and breakage of engineered pathways confounds consistent usage (Renda, Hammerling, and Barrick 2014). Our study employs experimental evolution as a tool for engineering, but as a method both to design and to test new strains. We utilized evolution experiments as a means both to discover the genetic underpinnings of a complex trait with real-world applications, and to determine and eliminate the most successful adaptive route in order to generate a more amenable strain background for future experiments. This approach represents a promising engineering technique not just for flocculation and related traits but also for traits such as antimicrobial resistance that represent major challenges of our time.

3.4 MATERIALS AND METHODS

3.4.1 *Strains and media used in this study*

The ancestral strain for all evolved strains used in this study was *Saccharomyces cerevisiae* laboratory haploid *MAT***a** strain FY4 (S288C), and backcrossing experiments were conducted using its isogenic *MAT*α counterpart FY5. Standard growth medium for overnight liquid cultures and agar plates used in this study was yeast extract peptone dextrose (YEPD) media, with 2% glucose and 1.7% agar for plates. Glucose-limited, sulfate-limited, and phosphate-limited liquid media and plates were prepared as in Gresham *et al* (Gresham et al. 2008) and detailed media recipes are available at http://dunham.gs.washington.edu/protocols.shtml.

To construct a *flo1* knockout strain, KanMX was amplified from the *FLO1* locus in the *flo1* strain from the yeast knockout collection (Giaever et al. 2002) using primers CJA009F/R

(Table 3.3). The PCR reaction was cleaned using a Zymo DNA Clean and Concentrator kit and DNA concentration was quantified with a Qubit fluorometer. Strain FY4 (S288C) was transformed with 1ug of the amplicon in 75µl of 1-step buffer (50% PEG4000 (40% final), 2M LioAc (0.2M final), 1M DTT (100nM final), salmon sperm carrier DNA) at 42C, and transformants were selected for G418 resistance. The *flo1::KanMX* strain was verified using Sanger sequencing.

3.4.2 *Multiplexed chemostat evolution experiments*

The first set of evolved clones was generated from 96 evolution experiments, conducted with laboratory strain FY4. The experiments were split equally between three nutrient limited conditions, 32 each of glucose, sulfate, and phosphate limitation, and organized into six blocks of 16 vessels maintained at 30°C. The evolution experiments were set up and media was prepared according to (Miller et al. 2013), with minor modifications. Sampling was conducted daily. The dilution rate was maintained in a target pump setting range of 0.16 and 0.18 volumes/hour, and generations elapsed were calculated as (1.44)*(time elapsed)*(dilution rate). Total generations were calculated as the cumulative sum of these individual times. One vessel was lost to pinched pump tubing that obstructed its media supply, for a final number of 95 evolution experiments. The remaining 95 evolution experiments were terminated at ~300 generations. Throughout the experiment, vessels were monitored for evidence of wall sticking and aggregation, and in this initial experiment both traits were scored together. In later experiments, we scored these traits separately. 12/32 phosphate-limited, 18/32 glucose-limited, and 3/31 sulfate-limited populations demonstrated evidence of aggregation or wall sticking, and we selected 9 phosphate, 11 glucose, and 3 sulfate-limited populations for further analysis.

The comparison between *flo1* knockout and wild-type strains was conducted using 64 glucose-limited chemostats run as above. Within each 16-vessel block, wild-type strains and knockout cultures were set up in alternating rows of 4. Up to 150 generations, sampling was conducted once weekly. Cultures were monitored daily for evidence of contamination, flocculation, and colonization in any of the media or effluent lines. After 150 generations, samples were stored twice weekly, and microscopy images for all cultures were saved once weekly. At the final timepoint, microscopy images were collected on all cultures. Clumps from the bottom of the culture or rings adhering to the vessel walls were collected with long sterile cotton swabs and resuspended in media and glycerol for storage. The final populations were plated on YPD to check for contamination and replica-plated onto G418 to validate the presence of the *KanMX* marker in only the expected *flo1* knockout populations.

3.4.3 *Clone isolation*

Colonies were struck out from glycerol stocks of the final time point of each experiment, inoculated into liquid culture and grown overnight at 30°C. From overnight cultures that displayed a clumping and/or settling phenotype, single cells were isolated using micromanipulation on a Nikon Eclipse 50i dissecting microscope, allowed to grow into colonies, screened for the phenotype in an overnight liquid culture of the appropriate nutrient-limited media, and saved at -80°C in glycerol stocks.

3.4.4 *Whole Genome Sequence analysis*

Genomic DNA for each clone was extracted using a Zymo YeaStar genomic DNA kit, checked for quality using a NanoDrop ND-1000 spectrophotometer, and quantified using an Invitrogen Qubit Fluorometer. Genomic DNA libraries were prepared for Illumina sequencing using the

Nextera sample preparation kit (Illumina) and sequenced using 150bp paired-end reads on an Illumina HiSeq. Ancestral DNA was prepared using a modified Hoffman-Winston preparation (Hoffman and Winston 1987).

Average sequence coverage from WGS of the clones was 97x. The reads were aligned against the genome sequence of sacCer3 using Burrows-Wheeler Aligner version 0.7.3 (H. Li and Durbin 2009). PCR duplicates were marked using Samblaster version 0.1.22 (Faust and Hall 2014) and indels were realigned using GATK version 3.5 (McKenna et al. 2010). For SNV and small indel analysis, variants were called using the bcftools call command (H. Li and Durbin 2009). SNVs/indels were filtered for quality and read depth, and mutations unique to the evolved clones were identified, annotated with a custom Python script (Pashkova et al. 2013), and verified by visual examination with the Integrative Genomics Viewer (IGV) (Robinson et al. 2011). This analysis revealed an average of three high quality SNVs/indels per clone after filtering, with a maximum of 16. Complete sequencing data for all of these clones is available under NCBI BioProject PRJNA339148, BioSample accessions SAMN05729740-5729793. Structural variants were called using lumpy (version accessed on 20160706) (Layer et al. 2014), and copy number variants were called using DNAcopy (Seshan and Olshen 2015) on 1000bp windows of coverage across the genome.

The deletion in gene *MIT1* was validated in clones YMD2694 and YMD3102 using PCR (primers EH053PF/PR) (Table 3.3) and Sanger sequencing. Validation of other mutations is described below.

Table 3.3 Primers used in this study

PRIMER	SEQUENCE 5'-3'	AMPLICON SIZE	EXPERIMENT
EH030PF	CAATATGCAAGCTCCTGGCA	2.2kb	Amplifies *FLO11* repeats from S288C - matches Up776flo11 primer from Zara 2009
EH030PR	GCCAGGGTATTTGGATGATG	2.2kb	pair for EH030PF
EH045PF	GAATTGTGCGGACGTTCCTC	507bp	Amplifies *HSL7* around potential secondary modifier mutation for YMD2683
EH045PR	GTGGAGGCGCCAATATTAGC	507bp	pair for EH045PF
EH046PF	CTGGCAGCGCTACTATCTCA	676bp	Amplifies *IRA1* around potential secondary modifier mutation for YMD2683
EH046PR	GCATTCACACTCGACTGCTT	676bp	pair for EH046PF
EH047PF	GCTACCTGCAATTGCATCAC	533bp	Amplifies *VTS1* around potential secondary modifier mutation for YMD2683
EH047PR	GACCAGCATTAGGATGCGTA	533bp	pair for EH047PF
EH048PF	GAGAGGCCACTGAGAGAGTA	596bp	Amplifies *TCP1* around potential secondary modifier mutation for YMD2683
EH048PR	CAGAGTCAGCACCAATGATC	596bp	pair for EH048PF
CJA007F	TCCACGGAGACATACGTTTG	2.1kb/8.1kb	Amplifies promoter region of *FLO1* to identify Ty insertions; 2.1kb without Ty; 8.1kb with Ty
CJA007R	TGTCCTCCGACAGAACCTAG	2.1kb/8.1kb	pair for CJA007F
CJA009F	TATTCGGAAGGCATGATGTC	2.5kb	Validates correct insertion of *FLO1* into S288C genome for knockout construction
CJA009R	TAAGCGAACCACACTAGATC	2.5kb	pair for CJA009F
EH052PF	GCTCATCCTTATTCGGCTCC	391bp	Amplifies *HOG1* around potential secondary modifier mutation for YMD2690
EH052PR	GTATGGCCTGGTTACCGTAG	391bp	pair for EH052PF
EH053PF	CTACAGCTCCTTATCCGGTG	425bp	Amplifies *MIT1* around deletion to confirm breakpoints for YMD2694
EH053PR	ATTGTTCGCGTGACCCATAG	425bp	pair for EH053PF
EH054PF	ATCTTGTTCTGGATGAGGCC	6.4kb	Amplifies *YBLWTy2-1* from S288C genome
EH054PR	CAAGAGGGAGCCGCTATTTC	6.4kb	pair for EH054PF

3.4.5 *Microscopy and validation of separation defects*

Strains were grown overnight in 5mL YEPD liquid culture at 30°C. 5µl of culture was examined microscopically at 150X magnification and photographed using a Canon Powershot SD1200 IS digital camera. Images were scored for evidence of mother-daughter separation defects, which were identified in two of the clones, YMD2680 and YMD2689. To validate the separation defects, calcofluor white was added to $1x10^7$ cells at a concentration of 100µg/mL, pipetted to

mix, and incubated in the dark for 5 minutes or more. Cells were pelleted at 13200 rpm for 1 minute and the supernatant was removed. The pellet was then washed vigorously with 500µl water three times and re-suspended in 50µl water. Bud scars were visualized using a DAPI filter at 630X magnification.

To validate true flocculation in the remaining clones, the evolved clones and ancestral strain were inoculated into 100µl YEPD cultures in two replicates in a round-bottom 96-well plate and grown overnight at 30°C without agitation. Cultures were re-suspended by pipetting and 5µl of culture was examined microscopically at 150X magnification and imaged. Cells were pelleted and the supernatant removed by pipetting, and one replicate was re-suspended in 100µl water and the other in 100µl 4mM EDTA. Each replicate was pipetted ten times and then examined microscopically and imaged. After ~50 minutes, replicates were re-suspended again by pipetting five times, and examined microscopically and imaged again.

3.4.6 *Quantitative settling assay*

Settling analysis was conducted according to the protocol described in Hope and Dunham 2014 (Hope and Dunham 2014). Briefly, each evolved clone or backcross segregant was grown in 5mL YEPD for 20 hours at 30°C; strain YMD2691 and its segregants are slow growing so an additional replicate was completed for these segregants with 30 hours of growth. Each culture tube was vortexed and then placed over a black background to settle for 60 minutes, with photos taken of the settling culture at time zero immediately after vortexing and at time 60 after an hour of settling. Images were converted to black and white in Picasa version 3.9.141.306 and analyzed in ImageJ version 1.47v (Abramoff, Magalhães, and Ram 2004). The settling ratio (percent of tube cleared at 60 minutes) was calculated as in Hope and Dunham 2014 (Hope and Dunham

2014). Three replicate measurements were taken on each image of the evolved clones, and a single measurement was made for the segregants.

3.4.7 *Backcrossing and settling segregation patterns*

All clones except YMD2680 and YMD2689, which had separation defects, were backcrossed to strain FY5. An average of 16 full tetrads per cross were dissected, with additional dissections for crosses with clones YMD2678 (24 tetrads total) and YMD2697 (38 tetrads total). Segregants were inoculated into 100µl YEPD in round-bottom 96-well plates and grown overnight at 30°C without agitation. Plates were re-suspended by gentle pipetting and allowed to settle without agitation for 15 minutes, when they were photographed and scored for settling ability.

3.4.8 *Bulk Segregant Analysis*

Crosses for clones YMD2684, 2686, 2687, 2696, 2697, 2698, and 2699 were not utilized for BSA after this point as it was determined that they harbored the Ty insertion in the *FLO1* promoter; four strains that harbored this insertion (YMD2681, 2683, 2685, and 2690) were included in BSA to verify causality for the Ty insertion. The cross with clone YMD2701 was also not included for BSA because it did not segregate the settling trait 2:2. The nine remaining strains without a *FLO1* promoter Ty element insertion or separation defect were analyzed using BSA. Segregants were binned into two pools of cells according to phenotype (settling or non-settling). Cells were pelleted, washed once with 500µl water, transferred to a 2mL lock-top eppendorf tube, pelleted again, decanted, and frozen at -20°C until DNA extraction. Genomic DNA was extracted using a modified Hoffman-Winston preparation (Hoffman and Winston 1987). Sequencing libraries were prepared using Nextera library preparation protocols as described for the original clones.

To identify causal mutations, BSA pools were analyzed similarly to the evolved clones but filtered individually by sample. For each sample, mutations present in both the settling and non-settling pool were removed. Mutations present at an allele frequency of 1 were determined to be causal.

3.4.9 *Identification of Ty element insertion location and element type*

A Ty insertion in the promoter of *FLO1* was identified in 12 of the evolved flocculent clones by visual examination in IGV and split read analysis tool retroSeq (Keane, Wong, and Adams 2013). These insertions were verified as full-length using PCR with primers CJA007F/R (Table S4). In some cases, an exact breakpoint was determined using the program lumpy, but for other samples the Ty element insertion location was determined by visual examination in IGV. All insertions placed the Ty element in reverse orientation with respect to the *FLO1* gene, determined by manual analysis of the mapping orientation of split reads.

A 2.1kb region upstream of *FLO1* and including the start of the ORF was amplified using PCR with Phusion polymerase and primer pair CJA007F/R (Table 3.3) for each clone with a Ty insertion identified in WGS. The presence of a Ty element insertion leading to a 6kb expansion was verified on a 1% agarose gel. PCR verification of the insertion failed in three clones, YMD2681, 2683, and 2697. The PCR reactions were cleaned using a Zymo DNA Clean and Concentrator kit and eluted in 100µl of water. Ty1 contains two EcoRI sites not shared with Ty2, and Ty2 contains a unique BamHI site missing from Ty1; these features facilitate classification of Ty type by restriction digest. The cleaned amplicons were split into two restriction enzyme digest reactions, one with EcoRI and the other with BamHI (New England BioLabs). A model of the amplified region was created in sequence analysis software Ape ("ApE- A Plasmid Editor" 2016), with a Ty insertion in the middle of each hot spot insertion region: Ty insertions were

observed between 95 and 156bp and between 394 and 470bp upstream of the *FLO1* ORF, so for the close insertion model a Ty1 was added at 125bp from the ORF and for the far insertion model at 432bp from the ORF. Predicted cutting with EcoRI for the close insertion site yielded four bands at 208, 1408, 2344, and 4118bp, and for the far insertion site four bands at 208, 1408, 2651, and 3811bp. We observed the three longest bands as predicted on a 1% agarose gel following the restriction digests, with distinct size differences between the mid and high bands for clones with known close and far insertions; for all evolved clones successfully analyzed, the insertion was classified as a Ty1. Predicted banding patterns for cleavage with BamHI in the region were also consistent with Ty1 elements. As a positive control, a known Ty2 element was amplified from the S288C genome using primers EH054PF/PR (Table 3.3) and the banding patterns that would be present for a Ty2 element with BamHI and EcoRI digests were confirmed.

3.4.10 *Crosses to determine* FLO1 *dependence of mutations in* ROX3*,* CSE2*, and* MIT1

*MAT*α segregants of clones with mutations in *CSE2* (YMD2678), *ROX3* (YMD2691), and *MIT1* (YMD2694) were crossed to a *flo1* knockout strain to facilitate examination of the phenotype of the double mutant progeny, recorded based on the settling ratio of segregants in 16-18 tetrads per cross. Mating types of segregants were verified using a standard halo mating assay (protocol available at http://dunham.gs.washington.edu/protocols.shtml).

3.4.11 *Additional analyses for secondary modifiers in clones YMD2683 and YMD2690*

Two strong candidates for clones with secondary modifiers were YMD2683, with an elongated cell morphology, and YMD2690, with an expansion of the internal repeats in *FLO11*. All of the segregants screened in the quantitative settling assay for clone YMD2683 were also tested for mutations in genes *HSL7, IRA1, VTS1*, and *TCP1* using primers EH045PF-EH048PR (Table 3.3)

and sent for Sanger sequencing by Genewiz. Microscopy was performed on all of the segregants from the YMD2683 cross and nine additional segregants were selected based on cell morphology (two with round suspended cells, two with long suspended cells, two with round flocculent cells, and three with long flocculent cells); all were analyzed with the quantitative settling assay and sequenced for mutations in *HSL7* and *IRA1*.

For seven settling segregants from the backcross with clone YMD2690, the *FLO11* internal repeat region was amplified using primers EH030PF/PR (Table 3.3) and results were examined on a 1% agarose gel. For the same segregants, the region of *HOG1* surrounding a premature stop in the clone was amplified using primers EH052PF/PR (Table 3.3) and sent for Sanger sequencing.

3.5 ACKNOWLEDGEMENTS

We thank Noah Hanson, Monica Sanchez, Erica Alcantara, Michelle Hays, Bryony Lynch, Mei Huang, and Annie Young for experimental assistance. We also thank Aimée Dudley and Matthew Bryce Taylor for helpful comments on the manuscript, students participating in the Cold Spring Harbor Laboratories Yeast Genetics and Genomics Course in 2014, 2015, and 2016 for their contributions to the Bulk Segregant Analysis components of this project, and Maxwell W. Libbrecht for manuscript review and statistics consultation. This project was supported by NSF grant 1120425 and NIH grant P41GM103533. This material is based in part upon work supported by the National Science Foundation under Cooperative Agreement No. DBI-0939454. The CSHL Yeast Course is supported by NSF grant MCB-1437145. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. CJA and CSH were supported

by T32 HG00035. MJD is a Rita Allen Foundation Scholar and a Senior Fellow in the Genetic Networks program at the Canadian Institute for Advanced Research.

Chapter 4. CONCLUDING REMARKS

Biofilm-related traits in yeast are desirable targets for genetic analysis, for their roles in both human health and industry. They are also models for the types of complex microbial traits that are particularly beneficial to explore using new methods. The work presented here demonstrates how studying biofilm-related traits in natural isolates can reveal the full spectrum of phenotypic diversity and the influence of genetic background, including ploidy, on phenotypic outcome. It also validates experimental evolution as a tool for determining the most successful adaptive routes for evolving complex traits, and which genes might make the greatest impact in strain engineering. At the intersection of these two projects, there is significant promise for the role of natural isolates with different adaptive potential for evolution-guided engineering.

4.1 DIVERSITY AND PLOIDY-DEPENDENCE OF BIOFILM-RELATED PHENOTYPES DEMONSTRATES VALUE OF WORKING WITH NATURAL ISOLATES

Prior to this work, very little was known about the prevalence of biofilm-related phenotypes in natural isolates of yeast. Many of these phenotypes are hypothesized to have important roles in the natural history of yeast, its ability to colonize new and beneficial environmental niches, and its ability to survive harsh conditions including antimicrobial treatments. Yeast has been isolated from a wide variety of sources, including from fermented beverages and foods, insect vectors, and even extreme high salt, high sugar, or subglacial environments (Liti et al. 2009; Breuer and Harms 2006; Turk et al. 2007; Thomas and Dieckmann 2002). It also has the potential to adapt

quickly genetically to these new environments, exemplified by the role of prions in natural isolates (Halfmann et al. 2012) and the use of transposable elements as an adaptive mechanism (Garfinkel 2005; Wilke and Adams 1992) (and demonstrated in Chapter 3). Although biofilms were suspected to be enriched in, or even the preferred phenotype for, natural isolates of yeast, the vast majority of experimental work with biofilm phenotypes has been conducted using biofilm-competent laboratory strains. A notable exception is the study of complex colony formation, which has been conducted largely in natural isolates and was an early indicator of the breadth of phenotypic diversity that might be available.

The phenotypic panels we present in Chapter 2 provide a comprehensive look at the variety of phenotypes achieved by diverse natural isolates. The quantitative assays we adapted and developed to assess relative amounts of flocculation, invasive growth, and surface adhesion are replicable and sensitive and importantly allow for direct comparison between phenotypes. There are many known differences between the traits, including key genetic contributors, nutrient responsiveness, components of the extracellular matrix, and cell morphology switching. The level of correlation between the traits was an outstanding question that we have begun to answer. Importantly, the lack of correlation between the quantitative traits indicates that each phenotype is fundamentally distinct and provides independent information about the biofilm-forming ability of a given strain. For example, even if a strain is capable of forming filamentous cells to invade media or generate colony complexity, that is not an indicator that both of those phenotypes will be observed, affirming that the purpose and underlying genetic complexity of these traits is likely different. The variation in trait strength and complexity depending on the strain background was also striking, and not dependent on the geographical or ecological niche of origin, or the genetic similarity of strains, a finding supported in evaluation of other traits in

natural isolates as well (Zörgö et al. 2013). Each of these findings demonstrates the importance of studying traits observed in natural isolates in the context of natural isolates themselves, and that generalizations made from work with laboratory strains will not necessarily apply.

A strong example of this is the variable relationship between ploidy and phenotype in natural isolates, one of the central findings in Chapter 2. Studies with laboratory strains have consistently demonstrated that biofilm-related phenotypes are weaker in diploid compared to haploid strains, and weaker still with further increases in ploidy (Reynolds and Fink 2001; Galitski et al. 1999). In other phenotypes unrelated to biofilm formation, results have been conflicting or even trended the opposite way. In a study of fermentation kinetics and CO_2 production rates, a 2N to 4N ploidy series of an industrial strain revealed higher CO_2 production rates with increasing ploidy, a contrast to previous studies, and a direct result of plasma membrane surface area for this particular phenotype (Salmon 1997), a variable that ploidy models often fail to include (Zörgö et al. 2013). Studies like this one, which used an industrial strain cultured in simulated enological conditions, suggest that the effect of ploidy might depend on strain background, phenotype, and growth conditions, similarly to the strain-dependent effects of ploidy we describe in Chapter 2 with the agar invasion phenotype in particular. Other work with natural isolates has further demonstrated important interactions between ploidy and the environment that help determine the phenotype produced by natural isolates, with no universally favored ploidy state (Zörgö et al. 2013).

Due to the additional phenotypes that may be accessible by working with natural isolates of varying ploidy, diploids are also an additional reservoir for adaptive potential. Polyploidy can drive adaptation through genome instability, with an increased rate of beneficial mutation and buffering of deleterious mutations that might later provide an adaptive benefit in a new

environment (Selmecki et al. 2015). The use of diploid natural isolates could yield significantly different results from their haploid counterparts in evolution experiments and for industrial applications.

4.2 EXPLOITING NATURAL VARIATION TO DETERMINE MECHANISMS UNDERLYING BIOFILM-RELATED TRAITS

The comprehensive panel of phenotypic data we have collected from natural isolates paves the way for additional experiments and analyses to determine the contributions of variation at the level of single genes and the mechanisms underlying the traits in different backgrounds. The extensive genetic variation harbored in these strains also creates opportunities for shuffling that variation in genes and regulatory regions of interest, observing the phenotypic outcome, and resolving Quantitative Trait Loci to Quantitative Trait Nucleotides.

4.2.1 *Determining the impact of variation in known biofilm-related genes on phenotype*

Of the many genes necessary for multiple biofilm-related traits, one of the most interesting cases is that of transcription factor *FLO8*. While other *FLO* genes might contribute to some phenotypes and not others depending on strain background, a functional *FLO8* regulates expression of all of them and is required unless their expression is induced (Smukalla et al. 2008). The impact on biofilm-related phenotypes from variation in *FLO1* and *FLO11* at the level of internal repeat expansion is known, but the influence of variation at the level of the upstream regulator for all of the traits is an open question and may be significant. The contribution of transcription factor variation to phenotypic variation is unknown, with studies both supporting and dismissing its role. Four transcription factors identified as QTNs acting on sporulation efficiency in wild strains were shown by Sudarsanam and Cohen to explain 98% of the

phenotypic variation recorded in sporulation efficiency, and activated different gene expression responses depending on the nature of the mutation in the TF (Sudarsanam and Cohen 2014). By contrast, Yvert *et al* show no enrichment in transcription factors among *trans* eQTLs in a mapping study designed to analyze genes involved in regulatory variation (Yvert et al. 2003). The involvement of *FLO8* in all biofilm-related traits and the detailed quantitative data available across highly genetically divergent natural isolates makes this a strong test scenario for investigating the question of the role in transcription factor variation in these and potentially other phenotypes.

To examine the role of *FLO8* specifically in biofilm-related phenotypes, I propose to create a collection of strains that differ only at the *FLO8* locus, using divergent alleles of *FLO8* from the Liti collection we characterize in Chapter 2, in a fixed background. With the phenotypes of each natural isolate now known, the contribution to those original phenotypes driven by variation in *FLO8* alone can be determined. "Congenic" strains that vary at a single locus have been created before by breeding (Lander and Schork 1994) and provide a platform for identifying polymorphic sites that do contribute to complex traits. Allelic collections generated from natural isolates have been already been used to explore the phenotypic effect of alleles underlying traits considered Mendelian and revealed the incredible background dependence of those genotypes (Hou et al. 2016).

The function of *FLO8* is also known to be affected by genetic background. One of the earliest studies demonstrating the involvement of *FLO8* in flocculation and invasive growth observed that phenotypes resulting from a working *FLO8* from Σ1278B in the context of a *flo8* S288C background were not quite as strong as the native Σ1278B phenotypes (Liu, Styles, and Fink 1996). Observing the phenotypic changes resulting from testing a collection of genetically

divergent *FLO8* alleles across multiple background strains with different native biofilm-related trait profiles could illuminate differences in the mechanism of action behind the phenotypes in each strain, especially coupled with analyzing changes in gene expression.

Approaching quantitative traits in this way is a fundamentally different method of quantitative trait analysis than QTL mapping. The growing number of natural isolates available to the yeast community does not lend itself well to the pairwise approach of traditional QTL studies. Instead of using QTL mapping to look broadly at variation across a genome in a single cross, this approach proposes to look deeply at variation in a single gene across many available natural isolates, enabling us to ask targeted questions about the role of many alleles of individual genes in specific traits of interest. The careful analysis of the impact of variation in a single gene in a fixed background context has the potential to yield insights about the contributions of individual alleles to complex traits, and if such analysis proves useful it could easily extend beyond *FLO8* and even beyond biofilm-related traits. This approach could be used to evaluate, for example, the flocculation profiles of different naturally occurring alleles of the genes identified in Chapter 3 for specific abilities. Contributions of small effect QTLs could be evaluated at the level of single alleles, and alleles with larger effect identified. With a phenotype more amenable to pooled experiments, like fitness, all of the alleles of a single gene from a large collection of natural isolates could be compared simultaneously.

4.2.2 *Identifying contributing genes using computational modules*

Although genes involved in biofilm-related traits have been identified in numerous mutagenesis and deletion collection studies, such studies are typically completed by exhaustive analysis of a single strain background and depend on a carefully created collection of deletion strains that cannot be reasonably duplicated for multiple backgrounds. As a result, much of what we know

about the genes potentially involved in biofilm-related traits comes from the analysis of a single laboratory strain. An alternative approach correlating gene expression modules with individual biofilm-related phenotypes would identify candidate genes and co-clustered novel genes in natural isolates. One such approach, called "INSPIRE" from Celik *et al* and applied successfully to ovarian cancer data, leverages existing gene expression datasets to create a low-dimensional representation of the data associated with the occurrence of a biological trait and identify potentially novel associated genes (Celik et al. 2016). Gene expression data for nearly two dozen of the Liti natural isolates is available from the "phenomics" study exploring phenotypes and gene expression in the collection (Skelly et al. 2013), providing a useful starting point to generate modules associated with biofilm-related traits; these modules could also be regenerated with gene expression data collected from yeast participating in biofilm-related phenotypes rather than suspended planktonically. This approach is different from typical coexpression networks which create modules of genes based on shared expression characteristics (Mackay, Stone, and Ayroles 2009) rather than the correlation between gene expression and quantitative phenotype data, as we would using INSPIRE.

4.2.3 *Identifying correlations between biofilm-related traits and stress response*

With increasing numbers of natural isolates under analysis by multiple groups, several types of data are now available for shared collections of strains. Studies by Kvitek *et al* and Warringer *et al* have assessed the phenotypes of natural isolates across hundreds of morphological criteria and environmental conditions, including chemical stresses (Kvitek, Will, and Gasch 2008; Warringer et al. 2011). Clustering of biofilm-related phenotypic data against stress resistance could reveal which phenotypes or combination of phenotypes are the most successful for resistance to multiple types of stress.

4.2.4 *Using "shuffling" methods to identify QTNs and examine regulatory variation*

One of the greatest remaining challenges in the mapping of complex traits is determining not just what loci (QTLs) and, with additional refinement and analysis, genes are contributing to a trait, but exactly which combination of nucleotides in a gene contribute to a trait: which, of the many variant sites, are the ones that are important. QTL mapping studies rely on recombination and high numbers of progeny to eventually sort background variants from mutations that co-segregate with the trait of interest. PCR shuffling refines this approach to forcibly recombine the genetic material of a single gene and so further identify which nucleotide or combination of nucleotides co-segregates with a trait. First described by "DNA shuffling" by Stemmer (Stemmer 1994) and improved by Zhao and Arnold, homologous genes can be fragmented and recombined into chimeric genes *in vitro* with a low rate of mutagenesis, combining only the variants present in the provided alleles (Zhao and Arnold 1997). The genetic diversity of natural isolates provides an opportunity to combine alleles from backgrounds with significantly different phenotypes, as in a QTL mapping study, and associate causal mutations with biofilm-related traits. It also allows for the possibility of creating genetic variants that drive stronger, weaker, or different phenotypes, with notable potential for industrial engineering purposes.

The advantage of a shuffling method over a more systematic overview of the impact of each individual possible variant observed using random mutagenesis is the ability to preserve and combinatorially test larger haplotype blocks and the effects of many naturally occurring variants together. We demonstrated in Chapter 3 that not all genes that have been associated with a trait in the lab represent successful evolutionary routes; similarly, not all variants of a gene we could construct in the lab would represent favorable variants in the wild. A shuffling method takes advantage of countless generations of adaptation to specific niches and enriches for

variants that are more likely to have a true adaptive advantage. The study of how variation in a transcription factor like *FLO8* affects downstream phenotypes would be enhanced by this type of analysis. It would also facilitate looking at variation in the *FLO8* regulatory region, ultimately combined with different alleles and tested in the context of different natural isolate backgrounds.

4.2.5 *Identifying natural isolates with the greatest engineering utility*

By analyzing the performance of different alleles of *FLO8* across different backgrounds, and shuffling *FLO8,* its promoter, or other biofilm-related genes, there are extensive possibilities for generating novel strains and phenotypes highly useful for engineering purposes. There are also opportunities to identify which of the existing natural isolates are most well adapted to each phenotype and the engineering potential they already possess.

In Chapter 2 we determine that not only are biofilm-related traits different in diploid and haploid natural isolates, but also that a few diploid strains deviate from the expectations of weaker phenotypes. Additional clustering analyses could be performed between haploids and diploids to determine which diploid strains are most similar to their haploid counterparts, more similar to other diploids, and which represent potentially useful outliers. Among the existing haploid and diploid strains, it is also unknown which are the most competitively successful across each trait. Recent innovations in isolating and counting cells from complex colonies and agar invasion (Helen Murphy, complex colonies, unpublished; Michael Dorrity, invasion, unpublished) might for the first time facilitate head-to-head or pooled competitions in the context of the relevant phenotype to identify the fittest isolates by using barcoded versions of the strains (Cubillos, Louis, and Liti 2009) to determine the proportion of each strain present at a final time point. This type of analysis could easily be extended to the *FLO8* allele strain collection or any shuffled variants of interest.

Finally, there are aspects of biofilm-related phenotypes in yeast that have directly to do with fitness; one of the most important applications of this is survival and reproduction under stressful conditions, including chemical treatment. A method using natural isolates in pairwise or pooled competition experiments could be extended into a continuous culture device with fitness as the trait of interest by using natural isolates to directly study the evolution of resistance to treatment with antimicrobial compounds. This could be further extended to test the performance of allele collections under dynamically changing drug concentrations, as in a morbidostat (Toprak et al. 2013).

4.3 Evolution experiments to reduce flocculation suggest a new approach to engineering

In synthetic biology, "evolutionary engineering" is frequently used to evolve strains that perform better under a particular selective pressure, most often relevant to industrial processes, e.g: high ethanol, high toxicity, changes in carbon source, and other environmental perturbations (Winkler and Kao 2014; Sonderegger and Sauer 2003; Çakar et al. 2005). Although the successful strains are often subjected to genetic and genomic analyses to determine the suite of mutations that led to the improvement, engineering using this method typically involves continued evolution of the most successful strains (Sonderegger and Sauer 2003; Çakar et al. 2005). By contrast, in Chapter 3 we use the understanding we gain of the genetic basis of an evolved trait to make a targeted strain design decision, and then test that strain using evolution again. This work combines two hallmarks of strain engineering, evolution and design, in a novel way and represents a new philosophy for how to explore biological systems.

4.3.1 *Early trends in evolution-guided design*

Recent trends in evolutionary engineering are moving away from the previous model of evolving strains with little or no performance in the condition of interest (e.g. xylose fermentation) under sometimes increasing selective pressure until a better performing strain is obtained (Sonderegger and Sauer 2003). Instead, new models incorporate an element of "rational design", an attempt to engineer a more favorable starting strain going into the evolution experiment. Chica *et al* describe the efficacy of rational design to improve enzyme efficacy, by targeting residues known to be important for binding specificity for random mutagenesis before evolution (Chica, Doucet, and Pelletier 2005). The approach we describe in Chapter 3, using the results of an evolution experiment to rationally design a new starting strain for subsequent evolution combines traditional evolutionary engineering and rational design methods in an entirely novel way. Importantly, the strain design choices are informed not by previous work in QTL mapping or deletion collection studies to identify candidate genes, but instead by the favored adaptive routes specific to evolution of the trait.

4.3.2 *Transposable elements as an adaptive reservoir*

As early as 1983, transposable elements were identified as contributors to increased adaptability by increasing the mutation rate in evolved strains of *Escherichia coli* (Chao et al. 1983), and a recent analysis of *E. coli* strains evolving for 50,000 generations also recorded a role for transposable Insertion Sequence elements in hypermutator strains (Tenaillon et al. 2016). In yeast, Wilke and Adams recorded instances of positive fitness adaptations driven by Ty1 insertions (Wilke and Adams 1992), and in *Drosophila simulans,* a transposable element insertion in a wild population of flies drives increased expression of a gene known to confer pesticide resistance (Schlenke and Begun 2004). Most recently, a study of the classic pigment

adaptation in peppered moth *Biston betularia,* which switched their wing color from pale to black during the Industrial Revolution, revealed a transposable element insertion increasing expression of a locus involved in wing patterning (Hof et al. 2016). Our study of the adaptive mechanisms underlying the evolution of flocculation in yeast contributes another piece to the body of evidence that Ty elements, far from being only selfish genetic elements, are an important genomic reservoir for rapid adaptation to new environments.

4.4 AGGREGATING CLONES AND POPULATIONS ARE A RESOURCE FOR FUTURE FLOCCULATION AND EVOLUTION EXPERIMENTS

In addition to answering fundamental questions about the utility of different adaptive routes to achieve aggregation, the evolved clones we study in Chapter 3 and their original underlying populations provide many opportunities for further study of flocculation and the role of transposable elements in evolution.

4.4.1 *Population dynamics and background effects on flocculation*

For each of the aggregating clones we analyze in Chapter 3, we have a frozen record of multiple time points throughout the evolution experiments as well as population-level sequencing data for one time point and the potential to generate more. By analyzing the frequency of the causal alleles we identified in population-level sequencing data from earlier time points, we could discover how early the causal mutations arose, when they occurred with respect to secondary modifier mutations, and determine whether there are any similarities in the types and timing of mutations across evolutions, including pathways that acquire mutations at particular times. An interesting open question is also whether the *FLO1* Ty insertion we observed most frequently can occur in any background or whether it can only happen in the context of a particular background.

This type of population-level analysis would reveal genetic interactions contributing to the evolution of aggregation and population dynamics, including the population-level allele frequency various types of aggregation mutations must reach before their phenotypic effects are observable.

Analysis of allele frequencies and the dynamics of the many mutations we observe in each clone over time might also help identify mutations associated with adaptation to the nutrient limitation and fitness, rather than aggregation. One interesting example of where we might observe this dynamic is in a flocculent clone that harbored both a Ty element insertion in the promoter of *FLO1* and a Chromosome I amplification duplicating the insertion and the gene. The resultant copy number increase did not enhance the flocculation of ability of the strain, at least as measured by our setting assay, leading us to hypothesize that the aneuploidy provides a different benefit. This particular clone was isolated from a phosphate-limited population, and previous work from Sunshine *et al* has recorded a modest fitness benefit for strains with a Chromosome I duplication under phosphate-limited growth (Sunshine et al. 2015); this represents just one of many mutations we recorded that might be an adaptive response to other aspects of the experiments.

4.4.2 *Flocculation as a thresholded trait*

In addition to asking what population frequency flocculation mutations must reach to be observed, it is also unknown to what extent flocculation itself is a threshold trait. Gene expression analysis of evolved strains with a *FLO1* Ty insertion but different flocculation levels could further reveal the role of *FLO1* expression on intensity of flocculation. Creating a collection of strains with different *FLO1* expression profiles, potentially by replacing or

shuffling its promoter, could reveal whether there is an expression threshold at which flocculation occurs, and the extent to which flocculation scales with *FLO1* expression.

4.4.3 *Untangling secondary modifiers of the flocculation trait*

Many of the flocculent clones harboring Ty insertions in the promoter of *FLO1* in Chapter 3 also show evidence of quantitative variation in flocculation among segregants and a possible role for one or several modifiers of flocculation specifically. For strains with clearly segregating alternative phenotypes, for example unusual cell morphology, additional BSA from pooling by these alternative phenotypes could serve to resolve causal alleles. For strains with quantitative variation among segregants but no clear secondary contributing phenotype, systematic complementation of known mutations with the wild-type allele or targeted sequencing of candidate alleles in segregants with a settling profile most divergent from the parent clone might reveal modifiers.

4.4.4 *Transposable element-driven adaptation: dependence on element type and mating type*

The flocculent clones we describe in Chapter 3 favor Ty element insertion in front of *FLO1* specifically. Restriction digest analysis of the inserted Ty elements revealed that they were all of type Ty1, which is the most abundant class of Ty elements in the *S. cerevisiae* genome. An interesting open question about this adaptive mechanism is whether Ty1 is specifically the only class of Ty element that can activate *FLO1* expression or whether it is simply the most likely class of element to mobilize by probability. This could be addressed by replacement of the *FLO1* Ty insertion with different classes of Ty element, at both the close and far insertion hotspot sites, to elucidate the role of Ty class in this particular adaptive route. Since we observe a Ty insertion in front of *FLO9* in one of our clones, it would also be promising to explore the efficacy of this

adaptive route in general by inserting Ty elements in front of the other *FLO* genes involved in cell-cell adhesion.

The role of Ty elements in adaptation in this system also raises questions about the role of mating type in this evolutionary route. The activity of Ty elements is known to be dependent on homozygosity at the mating type locus: while they are active in both haploid and a/a or **α**/**α** diploids, heterozygous diploids show reduced or abolished activity (Rothstein and Sherman 1980; Errede, Company, and Hutchison 1987; Fulton et al. 1988). Given the possible differences in Ty activation in haploid versus diploid backgrounds, as well as the ploidy-regulated differences we observed in flocculation ability among natural isolates, it would be useful to determine whether the adaptive routes we discovered would even be valid in a diploid evolution. It is possible that the same evolution experiment with a diploid strain would yield completely different results.

4.4.5 *Evolution of flocculation under varying transposable element activity*

Different species of yeast carry different transposable element burdens, as do different natural isolates of the same species (Kvitek, Will, and Gasch 2008), providing an avenue to explore the adaptive routes that lead to flocculation when the primary adaptive mechanism is enhanced or disabled. The same questions could be asked in experiments with additional treatments designed to increase or decrease the activity of Ty elements. The promoter of Ty1 is regulated by eight or more transcription factors, including TFs that cause invasive growth in response to nitrogen starvation (Todeschini et al. 2005; Morillon, Springer, and Lesage 2000). Nitrogen starvation as well as other stress conditions like exposure to DNA damaging agents cause increased activity of Ty1 (Todeschini et al. 2005), suggesting a potential relationship between Ty activation, stress response, and adaptation that leads to stress-protectant biofilm-related phenotypes.

4.5 Experimental evolution and natural isolates are both tools for engineering

The work we describe in these projects contributes to a larger goal of combining modern genetic and genomic techniques for efficient, rational engineering of strains with desirable properties. Understanding the contributions of variation in known biofilm-related genes to each of the biofilm-related phenotypes we describe in Chapter 2 could contribute to effective engineering of strains with very specific biofilm-related traits. The quantitative assays we have developed will provide ways to detect intermediate phenotypes and determine which variants drive biofilm-related phenotypes undesirable in medical environments, and other variants desirable for specific industrial uses, including brewing and biofuel production. In hospital environments, genotyping of specific loci could reveal whether a particular strain is likely to have increased virulence. If specific variants are identified that contribute to more damaging biofilm-related phenotypes, those variants also represent potential targets for small molecules or drugs. The isolates we study and the gene variants they provide are also a potential starting platform for other genome engineering techniques, including genome shuffling and directed evolution, which may further improve strains for industrial uses (Steensels and Verstrepen 2014). In-depth study of the gene variants in natural isolates contributing to specific biofilm-related traits will significantly advance this nascent field of strain engineering and generate generalizable techniques for other groups interested in using these diverse natural isolates for genome engineering projects. The phenotypic variation of natural isolates also includes industrially and medically relevant traits, like fermentation capacity, tolerance to chemical treatments, and flocculation ability (Wohlbach et al. 2014), and natural isolates with favorable phenotypes represent new potential chassis for industrial processes, and even novel starting points for evolution experiments.

BIBLIOGRAPHY

Aa, Erlend, Jeffrey P. Townsend, Rachel I. Adams, Kaare M. Nielsen, and John W. Taylor. 2006. "Population Structure and Gene Evolution in Saccharomyces Cerevisiae." *FEMS Yeast Research* 6 (5): 702–15. doi:10.1111/j.1567-1364.2006.00059.x.

Abramoff, M. D., Paulo J. Magalhães, and Sunanda J. Ram. 2004. "Image Processing with ImageJ." Article. *Biophotonics International*. http://dspace.library.uu.nl/handle/1874/204900.

Adams, J, and P W Oeller. 1986. "Structure of Evolving Populations of Saccharomyces Cerevisiae: Adaptive Changes Are Frequently Associated with Sequence Alterations Involving Mobile Elements Belonging to the Ty Family." *Proceedings of the National Academy of Sciences of the United States of America* 83 (18): 7124–27.

Adams, Julian. 2004. "Microbial Evolution in Laboratory Environments." *Research in Microbiology* 155 (March): 311–18. doi:10.1016/j.resmic.2004.01.013.

Anderson, Tim, Standwell Nkhoma, Andrea Ecker, and David Fidock. 2011. "How Can We Identify Parasite Genes That Underlie Antimalarial Drug Resistance?" *Pharmacogenomics* 12 (1): 59–85. doi:10.2217/pgs.10.165.

"ApE - A Plasmid Editor." 2016. Accessed September 15. http://biologylabs.utah.edu/jorgensen/wayned/ape/.

Araya, Carlos L., Celia Payen, Maitreya J. Dunham, and Stanley Fields. 2010. "Whole-Genome Sequencing of a Laboratory-Evolved Yeast Strain." *BMC Genomics* 11: 88. doi:10.1186/1471-2164-11-88.

Baillie, G. S., and L. J. Douglas. 2000. "Matrix Polymers of Candida Biofilms and Their Possible Role in Biofilm Resistance to Antifungal Agents." *The Journal of Antimicrobial Chemotherapy* 46 (3): 397–403.

Baillie, George S., and L. Julia Douglas. 1999. "Role of Dimorphism in the Development of Candida Albicans Biofilms." *Journal of Medical Microbiology* 48 (7): 671–79. doi:10.1099/00222615-48-7-671.

Barnett, James Arthur, Roger William Payne, and David Yarrow. 1983. *Yeasts : Characteristics and Identification*. Cambridge University Press.

Basehoar, Andrew D, Sara J Zanton, and B. Franklin Pugh. 2004. “Identification and Distinct Regulation of Yeast TATA Box-Containing Genes.” *Cell* 116 (5): 699–709. doi:10.1016/S0092-8674(04)00205-3.

Bauer, Florian F., Patrick Govender, and Michael C. Bester. 2010. “Yeast Flocculation and Its Biotechnological Relevance.” *Applied Microbiology and Biotechnology* 88 (1): 31–39. doi:10.1007/s00253-010-2783-0.

Bayly, Jennifer C., Lois M. Douglas, Isak S. Pretorius, Florian F. Bauer, and Anne M. Dranginis. 2005. “Characteristics of Flo11-Dependent Flocculation in Saccharomyces Cerevisiae.” *FEMS Yeast Research* 5 (12): 1151–56. doi:10.1016/j.femsyr.2005.05.004.

Beauvais, Anne, Céline Loussert, Marie Christine Prevost, Kevin Verstrepen, and Jean Paul Latgé. 2009. “Characterization of a Biofilm-like Extracellular Matrix in FLO1-Expressing Saccharomyces Cerevisiae Cells.” *FEMS Yeast Research* 9 (3): 411–19. doi:10.1111/j.1567-1364.2009.00482.x.

Bender, A, and J R Pringle. 1991. “Use of a Screen for Synthetic Lethal and Multicopy Suppressee Mutants to Identify Two New Genes Involved in Morphogenesis in Saccharomyces Cerevisiae.” *Molecular and Cellular Biology* 11 (3): 1295–1305.

Bergström, Anders, Jared T. Simpson, Francisco Salinas, Benjamin Barré, Leopold Parts, Amin Zia, Alex N. Nguyen Ba, et al. 2014. “A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes.” *Molecular Biology and Evolution* 31 (4): 872–88. doi:10.1093/molbev/msu037.

Beveridge, Tj. 2001. “Use of the Gram Stain in Microbiology.” *Biotechnic & Histochemistry* 76 (3): 111–18. doi:10.1080/bih.76.3.111.118.

Birkeland, Shanda R., Natsuko Jin, Alev Cagla Ozdemir, Robert H. Lyons, Lois S. Weisman, and Thomas E. Wilson. 2010. “Discovery of Mutations in Saccharomyces Cerevisiae by Pooled Linkage Analysis and Whole-Genome Sequencing.” *Genetics* 186 (4): 1127–37. doi:10.1534/genetics.110.123232.

Bleykasten-Grosshans, Claudine, Anne Friedrich, and Joseph Schacherer. 2013. “Genome-Wide Analysis of Intraspecific Transposon Diversity in Yeast.” *BMC Genomics* 14: 399. doi:10.1186/1471-2164-14-399.

Bloom, Joshua S., Ian M. Ehrenreich, Wesley T. Loo, Thúy-Lan Võ Lite, and Leonid Kruglyak. 2013. "Finding the Sources of Missing Heritability in a Yeast Cross." *Nature* 494 (7436): 234–37. doi:10.1038/nature11867.

Blum, K., E. P. Noble, P. J. Sheridan, A. Montgomery, T. Ritchie, P. Jagadeeswaran, H. Nogami, A. H. Briggs, and J. B. Cohn. 1990. "Allelic Association of Human Dopamine D2 Receptor Gene in Alcoholism." *JAMA* 263 (15): 2055–60.

"Boeke: 4 Yeast Transposable Elements." 2016. Accessed September 13. https://cshmonographs.org/index.php/monographs/article/view/3183/2523.

Bombieri, Cristina, Manuela Seia, and Carlo Castellani. 2015. "Genotypes and Phenotypes in Cystic Fibrosis and Cystic Fibrosis Transmembrane Regulator-Related Disorders." *Seminars in Respiratory and Critical Care Medicine* 36 (2): 180–93. doi:10.1055/s-0035-1547318.

Bony, Muriel, Pierre Barre, and Bruno Blondin. 1998. "Distribution of the Flocculation Protein, Flop, at the Cell Surface during Yeast Growth: The Availability of Flop Determines the Flocculation Level." *Yeast* 14 (1): 25–35. doi:10.1002/(SICI)1097-0061(19980115)14:1<25::AID-YEA197>3.0.CO;2-C.

Borneman, Anthony R., Justine A. Leigh-Bell, Haiyuan Yu, Paul Bertone, Mark Gerstein, and Michael Snyder. 2006. "Target Hub Proteins Serve as Master Regulators of Development in Yeast." *Genes & Development* 20 (4): 435–48. doi:10.1101/gad.1389306.

Botstein, D, R L White, M Skolnick, and R W Davis. 1980. "Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms." *American Journal of Human Genetics* 32 (3): 314–31.

Brauer, Matthew J., Cheryl M. Christianson, Dave A. Pai, and Maitreya J. Dunham. 2006. "Mapping Novel Traits by Array-Assisted Bulk Segregant Analysis in Saccharomyces Cerevisiae." *Genetics* 173 (3): 1813–16. doi:10.1534/genetics.106.057927.

Brem, R. B. 2002. "Genetic Dissection of Transcriptional Regulation in Budding Yeast." *Science* 296 (5568): 752–55. doi:10.1126/science.1069516.

Breuer, Uta, and Hauke Harms. 2006. "Debaryomyces Hansenii — an Extremophilic Yeast with Biotechnological Potential." *Yeast* 23 (6): 415–37. doi:10.1002/yea.1374.

Brown, C. J., K. M. Todd, and R. F. Rosenzweig. 1998. “Multiple Duplications of Yeast Hexose Transport Genes in Response to Selection in a Glucose-Limited Environment.” *Molecular Biology and Evolution* 15 (8): 931–42.

Brown, T A, C Evangelista, and B L Trumpower. 1995. “Regulation of Nuclear Genes Encoding Mitochondrial Proteins in *Saccharomyces cerevisiae*.” *Journal of Bacteriology* 177 (23): 6836–43.

Brückner, Stefan, and Hans-Ulrich Mösch. 2012. “Choosing the Right Lifestyle: Adhesion and Development in Saccharomyces Cerevisiae.” *FEMS Microbiology Reviews* 36 (1): 25–58. doi:10.1111/j.1574-6976.2011.00275.x.

Cain, Christopher W., Matthew B. Lohse, Oliver R. Homann, Anita Sil, and Alexander D. Johnson. 2012. “A Conserved Transcriptional Regulator Governs Fungal Morphology in Widely Diverged Species.” *Genetics* 190 (2): 511–21. doi:10.1534/genetics.111.134080.

Çakar, Z. Petek, Urartu O. S. Seker, Candan Tamerler, Marco Sonderegger, and Uwe Sauer. 2005. “Evolutionary Engineering of Multiple-Stress Resistant Saccharomyces Cerevisiae.” *FEMS Yeast Research* 5 (6–7): 569–78. doi:10.1016/j.femsyr.2004.10.010.

Carrico, Pauline M., and Richard S. Zitomer. 1998. “Mutational Analysis of the Tup1 General Repressor of Yeast.” *Genetics* 148 (2): 637–44.

Celik, Safiye, Benjamin A. Logsdon, Stephanie Battle, Charles W. Drescher, Mara Rendi, R. David Hawkins, and Su-In Lee. 2016. “Extracting a Low-Dimensional Description of Multiple Gene Expression Datasets Reveals a Potential Driver for Tumor-Associated Stroma in Ovarian Cancer.” *Genome Medicine* 8: 66. doi:10.1186/s13073-016-0319-7.

Chao, Lin, Christopher Vargas, Brian B. Spear, and Edward C. Cox. 1983. “Transposable Elements as Mutator Genes in Evolution.” *Nature* 303 (5918): 633–35. doi:10.1038/303633a0.

Chica, Roberto A, Nicolas Doucet, and Joelle N Pelletier. 2005. “Semi-Rational Approaches to Engineering Enzyme Activity: Combining the Benefits of Directed Evolution and Rational Design.” *Current Opinion in Biotechnology*, Protein technologies and commercial enzymes, 16 (4): 378–84. doi:10.1016/j.copbio.2005.06.004.

Clemons, Karl V., John H. McCusker, Ronald W. Davis, and David A. Stevens. 1994. “Comparative Pathogenesis of Clinical and Nonclinical Isolates of Saccharomyces

Cerevisiae." *Journal of Infectious Diseases* 169 (4): 859–67. doi:10.1093/infdis/169.4.859.

Cloninger, C. Robert. 1991. "D2 Dopamine Receptor Gene Is Associated but Not Linked With Alcoholism." *JAMA: The Journal of the American Medical Association* 266 (13): 1833. doi:10.1001/jama.1991.03470130113040.

Coney, L R, and G S Roeder. 1988. "Control of Yeast Gene Expression by Transposable Elements: Maximum Expression Requires a Functional Ty Activator Sequence and a Defective Ty Promoter." *Molecular and Cellular Biology* 8 (10): 4009–17.

Cormack, B. P., N. Ghori, and S. Falkow. 1999. "An Adhesin of the Yeast Pathogen Candida Glabrata Mediating Adherence to Human Epithelial Cells." *Science (New York, N.Y.)* 285 (5427): 578–82.

Cromie, Gareth A., Katie E. Hyma, Catherine L. Ludlow, Cecilia Garmendia-Torres, Teresa L. Gilbert, Patrick May, Angela A. Huang, Aimée M. Dudley, and Justin C. Fay. 2013. "Genomic Sequence Diversity and Population Structure of Saccharomyces Cerevisiae Assessed by RAD-Seq." *G3: Genes|Genomes|Genetics* 3 (12): 2163–71. doi:10.1534/g3.113.007492.

Cubillos, Francisco A., Eleonora Billi, Enikö Zörgö, Leopold Parts, Patrick Fargier, Stig Omholt, Anders Blomberg, Jonas Warringer, Edward J. Louis, and Gianni Liti. 2011. "Assessing the Complex Architecture of Polygenic Traits in Diverged Yeast Populations." *Molecular Ecology* 20 (7): 1401–13. doi:10.1111/j.1365-294X.2011.05005.x.

Cubillos, Francisco A., Edward J. Louis, and Gianni Liti. 2009. "Generation of a Large Set of Genetically Tractable Haploid and Diploid Saccharomyces Strains." *FEMS Yeast Research* 9 (8): 1217–25. doi:10.1111/j.1567-1364.2009.00583.x.

Cubillos, Francisco A., Leopold Parts, Francisco Salinas, Anders Bergström, Eugenio Scovacricchi, Amin Zia, Christopher J. R. Illingworth, et al. 2013. "High-Resolution Mapping of Complex Traits with a Four-Parent Advanced Intercross Yeast Population." *Genetics* 195 (3): 1141–55. doi:10.1534/genetics.113.155515.

Cullen, Paul J. 2015. "Evaluating the Activity of the Filamentous Growth MAPK Pathway in Yeast." *Cold Spring Harbor Protocols* 2015 (3): 276–83. doi:10.1101/pdb.prot085092.

Cullen, Paul J., and George F. Sprague. 2000. “Glucose Depletion Causes Haploid Invasive Growth in Yeast.” *Proceedings of the National Academy of Sciences* 97 (25): 13619–24. doi:10.1073/pnas.240345197.

Deutschbauer, Adam M., and Ronald W. Davis. 2005. “Quantitative trait loci mapped to single nucleotide resolution in yeast.” *Nature Genetics* 37 (12): 1333-40.

Dilda, Christy L., and Trudy F. C. Mackay. 2002. “The Genetic Architecture of Drosophila Sensory Bristle Number.” *Genetics* 162 (4): 1655–74.

Domingues, Lucília, António A. Vicente, Nelson Lima, and José A. Teixeira. n.d. “Applications of Yeast Flocculation in Biotechnological Processes.” *Biotechnology and Bioprocess Engineering* 5 (4): 288–305. doi:10.1007/BF02942185.

Doniger, Scott W., Hyun Seok Kim, Devjanee Swain, Daniella Corcuera, Morgan Williams, Shiaw-Pyng Yang, and Justin C. Fay. 2008. “A Catalog of Neutral and Deleterious Polymorphism in Yeast.” *PLOS Genet* 4 (8): e1000183. doi:10.1371/journal.pgen.1000183.

Donnelly, Peter. 2008. “Progress and Challenges in Genome-Wide Association Studies in Humans.” *Nature* 456 (7223): 728–31. doi:10.1038/nature07631.

Douglas, L. Julia. 2003. “Candida Biofilms and Their Role in Infection.” *Trends in Microbiology* 11 (1): 30–36.

Drees, Becky L., Vesteinn Thorsson, Gregory W. Carter, Alexander W. Rives, Marisa Z. Raymond, Iliana Avila-Campillo, Paul Shannon, and Timothy Galitski. 2005. “Derivation of Genetic Interaction Networks from Quantitative Phenotype Data.” *Genome Biology* 6: R38. doi:10.1186/gb-2005-6-4-r38.

Dunham, Maitreya J., Hassan Badrane, Tracy Ferea, Julian Adams, Patrick O. Brown, Frank Rosenzweig, and David Botstein. 2002. “Characteristic Genome Rearrangements in Experimental Evolution of Saccharomyces Cerevisiae.” *Proceedings of the National Academy of Sciences* 99 (25): 16144–49. doi:10.1073/pnas.242624799.

Dunn, Barbara, Chandra Richter, Daniel J. Kvitek, Tom Pugh, and Gavin Sherlock. 2012. “Analysis of the Saccharomyces Cerevisiae Pan-Genome Reveals a Pool of Copy Number Variants Distributed in Diverse Yeast Strains from Differing Industrial Environments.” *Genome Research* 22 (5): 908–24. doi:10.1101/gr.130310.111.

Ehrenreich, Ian M., Noorossadat Torabi, Yue Jia, Jonathan Kent, Stephen Martis, Joshua A. Shapiro, David Gresham, Amy A. Caudy, and Leonid Kruglyak. 2010. "Dissection of Genetically Complex Traits with Extremely Large Pools of Yeast Segregants." *Nature* 464 (7291): 1039–42. doi:10.1038/nature08923.

Ellis, Tom, Xiao Wang, and James J. Collins. 2009. "Diversity-Based, Model-Guided Construction of Synthetic Gene Networks with Predicted Functions." *Nature Biotechnology* 27 (5): 465–71. doi:10.1038/nbt.1536.

Errede, B, T S Cardillo, M A Teague, and F Sherman. 1984. "Identification of Regulatory Regions within the Ty1 Transposable Element That Regulate Iso-2-Cytochrome c Production in the CYC7-H2 Yeast Mutant." *Molecular and Cellular Biology* 4 (7): 1393–1401.

Errede, B., M. Company, and C. A. Hutchison. 1987. "Ty1 Sequence with Enhancer and Mating-Type-Dependent Regulatory Activities." *Molecular and Cellular Biology* 7 (1): 258–65. doi:10.1128/MCB.7.1.258.

Faust, Gregory G., and Ira M. Hall. 2014. "SAMBLASTER: Fast Duplicate Marking and Structural Variant Read Extraction." *Bioinformatics* 30 (17): 2503–5. doi:10.1093/bioinformatics/btu314.

Fay, Justin C., and Joseph A. Benavides. 2005. "Evidence for Domesticated and Wild Populations of Saccharomyces Cerevisiae." *PLOS Genet* 1 (1): e5. doi:10.1371/journal.pgen.0010005.

Fay, Justin C, Heather L McCullough, Paul D Sniegowski, and Michael B Eisen. 2004. "Population Genetic Variation in Gene Expression Is Associated with Phenotypic Variation in Saccharomyces Cerevisiae." *Genome Biology* 5 (4): R26.

Fekih-Salem, R., J. Harmand, C. Lobry, A. Rapaport, and T. Sari. 2013. "Extensions of the Chemostat Model with Flocculation." *Journal of Mathematical Analysis and Applications* 397 (1): 292–306. doi:10.1016/j.jmaa.2012.07.055.

Fichtner, Lars, Florian Schulze, and Gerhard H Braus. 2007. "Differential Flo8p-Dependent Regulation of FLO1 and FLO11 for Cell–cell and Cell–substrate Adherence of S. Cerevisiae S288c." *Molecular Microbiology* 66 (5): 1276–89. doi:10.1111/j.1365-2958.2007.06014.x.

Fidalgo, Manuel, Ramon R. Barrales, and Juan Jimenez. 2008. "Coding Repeat Instability in the *FLO11* Gene of *Saccharomyces* Yeasts." *Yeast* 25 (12): 879–89. doi:10.1002/yea.1642.

Fleming, Alastair B., Suzanne Beggs, Michael Church, Yoshihiro Tsukihashi, and Sari Pennings. 2014. "The Yeast Cyc8–Tup1 Complex Cooperates with Hda1p and Rpd3p Histone Deacetylases to Robustly Repress Transcription of the Subtelomeric FLO1 Gene." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839 (11): 1242–55. doi:10.1016/j.bbagrm.2014.07.022.

Flint, Jonathan, and Trudy F. C. Mackay. 2009. "Genetic Architecture of Quantitative Traits in Mice, Flies, and Humans." *Genome Research* 19 (5): 723–33. doi:10.1101/gr.086660.108.

Flint, Jonathan, and Richard Mott. 2001. "Finding the Molecular Basis of Quatitative Traits: Successes and Pitfalls." *Nature Reviews Genetics* 2 (6): 437–45. doi:10.1038/35076585.

Frazer, Kelly A., Sarah S. Murray, Nicholas J. Schork, and Eric J. Topol. 2009. "Human genetic variation and its contribution to complex traits." *Nature Reviews Genetics* 10 (4): 241-51.

Fujita, Atsushi, Akio Tonouchi, Takatoshi Hiroko, Fumika Inose, Takeyuki Nagashima, Rika Satoh, and Shigeko Tanaka. 1999. "Hsl7p, a Negative Regulator of Ste20p Protein Kinase in the Saccharomyces Cerevisiae Filamentous Growth-Signaling Pathway." *Proceedings of the National Academy of Sciences* 96 (15): 8522–27. doi:10.1073/pnas.96.15.8522.

Fulton, Alexandra M., Peter D. Rathjen, Susan M. Kingsman, and Alan J. Kingsman. 1988. "Upstream and Downstream Transcriptional Control Signals in the Yeast Retrotransposon, TY." *Nucleic Acids Research* 16 (12): 5439–58. doi:10.1093/nar/16.12.5439.

Galitski, Timothy, Alok J. Saldanha, Cora A. Styles, Eric S. Lander, and Gerald R. Fink. 1999. "Ploidy Regulation of Gene Expression." *Science* 285 (5425): 251–54. doi:10.1126/science.285.5425.251.

Garcia Sanchez, Rosa, Kaisa Karhumaa, César Fonseca, Violeta Sànchez Nogué, João RM Almeida, Christer U. Larsson, Oskar Bengtsson, Maurizio Bettiga, Bärbel Hahn-Hägerdal, and Marie F. Gorwa-Grauslund. 2010. "Improved Xylose and Arabinose Utilization by an Industrial Recombinant Saccharomyces Cerevisiae Strain Using Evolutionary Engineering." *Biotechnology for Biofuels* 3: 13. doi:10.1186/1754-6834-3-13.

Garfinkel, D.J. 2005. "Genome Evolution Mediated by Ty Elements in Saccharomyces." *Cytogenetic and Genome Research* 110 (1–4): 63–69. doi:10.1159/000084939.

Gelernter, J., D. Goldman, and N. Risch. 1993. "The A1 Allele at the D2 Dopamine Receptor Gene and Alcoholism. A Reappraisal." *JAMA* 269 (13): 1673–77.

Gerke, Justin P., Christina T. L. Chen, and Barak A. Cohen. 2006. "Natural Isolates of Saccharomyces Cerevisiae Display Complex Genetic Variation in Sporulation Efficiency." *Genetics* 174 (2): 985–97. doi:10.1534/genetics.106.058453.

Gerke, Justin, Kim Lorenz, and Barak Cohen. 2009. "Genetic interactions between transcription factors cause natural variation in yeast." *Science* 323 (5913): 498-501.

Gerstein, Aleeza C., Dara S. Lo, and Sarah P. Otto. 2012. "Parallel Genetic Changes and Nonparallel Gene–Environment Interactions Characterize the Evolution of Drug Resistance in Yeast." *Genetics* 192 (1): 241–52. doi:10.1534/genetics.112.142620.

Giaever, Guri, Angela M. Chu, Li Ni, Carla Connelly, Linda Riles, Steeve Véronneau, Sally Dow, et al. 2002. "Functional Profiling of the Saccharomyces Cerevisiae Genome." *Nature* 418 (6896): 387–91. doi:10.1038/nature00935.

Gimeno, Carlos J., Per O. Ljungdahl, Cora A. Styles, and Gerald R. Fink. 1992. "Unipolar Cell Divisions in the Yeast S. Cerevisiae Lead to Filamentous Growth: Regulation by Starvation and RAS." *Cell* 68 (6): 1077–90. doi:10.1016/0092-8674(92)90079-R.

Gould, Stephen Jay. 1990. *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton & Company.

Granek, Joshua A., and Paul M. Magwene. 2010. "Environmental and Genetic Determinants of Colony Morphology in Yeast." *PLOS Genet* 6 (1): e1000823. doi:10.1371/journal.pgen.1000823.

Granek, Joshua A., Debra Murray, Ömür Kayrkçi, and Paul M. Magwene. 2013. "The Genetic Architecture of Biofilm Formation in a Clinical Isolate of Saccharomyces Cerevisiae." *Genetics* 193 (2): 587–600. doi:10.1534/genetics.112.142067.

Gresham, David, Michael M. Desai, Cheryl M. Tucker, Harry T. Jenq, Dave A. Pai, Alexandra Ward, Christopher G. DeSevo, David Botstein, and Maitreya J. Dunham. 2008. "The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast." *PLOS Genet* 4 (12): e1000303. doi:10.1371/journal.pgen.1000303.

Gresham, David and Jungeui Hong. 2015. "The functional basis of adaptive evolution in chemostats." *FEMS Micrbiology Review* 39 (1): 2-16. doi:10.1111/1574-6976.12082

Guo, Bing, Cora A. Styles, Qinghua Feng, and Gerald R. Fink. 2000. "A Saccharomyces Gene Family Involved in Invasive Growth, Cell–cell Adhesion, and Mating." *Proceedings of the National Academy of Sciences* 97 (22): 12158–63. doi:10.1073/pnas.220420397.

Halfmann, Randal, Daniel F. Jarosz, Sandra K. Jones, Amelia Chang, Alex K. Lancaster, and Susan Lindquist. 2012. "Prions Are a Common Mechanism for Phenotypic Inheritance in Wild Yeasts." *Nature* 482 (7385): 363–68. doi:10.1038/nature10875.

Halme, Adrian, Stacie Bumgarner, Cora Styles, and Gerald R. Fink. 2004. "Genetic and Epigenetic Regulation of the FLO Gene Family Generates Cell-Surface Variation in Yeast." *Cell* 116 (3): 405–15. doi:10.1016/S0092-8674(04)00118-7.

Heil, Caiti S. Smukowski, Christopher G. DeSevo, Dave A. Pai, Cheryl M. Tucker, Margaret L. Hoang, and Maitreya J. Dunham. 2016. "Selection on Heterozygosity Drives Adaptation in Intra- and Interspecific Hybrids." *bioRxiv*, September, 73007. doi:10.1101/073007.

Hof, Arjen E. van't, Pascal Campagne, Daniel J. Rigden, Carl J. Yung, Jessica Lingley, Michael A. Quail, Neil Hall, Alistair C. Darby, and Ilik J. Saccheri. 2016. "The Industrial Melanism Mutation in British Peppered Moths Is a Transposable Element." *Nature* 534 (7605): 102–5. doi:10.1038/nature17951.

Hoffman, C. S., and F. Winston. 1987. "A Ten-Minute DNA Preparation from Yeast Efficiently Releases Autonomous Plasmids for Transformation of Escherichia Coli." *Gene* 57 (2–3): 267–72.

Holmes, Daniel L., Alex K. Lancaster, Susan Lindquist, and Randal Halfmann. 2013. "Heritable Remodeling of Yeast Multicellularity by an Environmentally Responsive Prion." *Cell* 153 (1): 153–65. doi:10.1016/j.cell.2013.02.026.

Hong, Jungeui, and David Gresham. 2014. "Molecular Specificity, Convergence and Constraint Shape Adaptive Evolution in Nutrient-Poor Environments." *PLoS Genetics* 10 (1). doi:10.1371/journal.pgen.1004041.

Hoon, M. J. L. de, S. Imoto, J. Nolan, and S. Miyano. 2004. "Open Source Clustering Software." *Bioinformatics* 20 (9): 1453–54. doi:10.1093/bioinformatics/bth078.

Hope, Elyse A., and Maitreya J. Dunham. 2014. "Ploidy-Regulated Variation in Biofilm-Related Phenotypes in Natural Isolates of Saccharomyces Cerevisiae." *G3: Genes|Genomes|Genetics* 4 (9): 1773–86. doi:10.1534/g3.114.013250.

Hou, Jing, Anastasie Sigwalt, David Pflieger, Jackson Peter, Jacky de Montigny, Maitreya Dunham, and Joseph Schacherer. 2016. "The hidden complexity of Mendelian traits across yeast natural populations." *Cell Reports* 16 (4): 1106-1114.

Huang, Zhiwei, Kaifu Chen, Jianhuai Zhang, Yongxiang Li, Hui Wang, Dandan Cui, Jiangwu Tang, et al. 2013. "A Functional Variomics Tool for Discovering Drug Resistance Genes and Drug Targets." *Cell Reports* 3 (2): 577–85. doi:10.1016/j.celrep.2013.01.019.

Jansen, Ritsert C., and Jan-Peter Nap. 2001. "Genetical Genomics: The Added Value from Segregation." *Trends in Genetics* 17 (7): 388–91. doi:10.1016/S0168-9525(01)02310-1.

Jara, Matías, Francisco A. Cubillos, Verónica García, Francisco Salinas, Omayra Aguilera, Gianni Liti, and Claudio Martínez. 2014. "Mapping Genetic Variants Underlying Differences in the Central Nitrogen Metabolism in Fermenter Yeasts." *PLOS ONE* 9 (1): e86533. doi:10.1371/journal.pone.0086533.

Jin, Rui, Craig J. Dobry, Phillip J. McCown, and Anuj Kumar. 2008. "Large-Scale Analysis of Yeast Filamentous Growth by Systematic Gene Disruption and Overexpression." *Molecular Biology of the Cell* 19 (1): 284–96. doi:10.1091/mbc.E07-05-0519.

Johannsen, W. 1911. "The Genotype Conception of Heredity." *The American Naturalist* 45 (531): 129–59.

Johnston, J. R., and H. P. Reader. 1983. “Genetic Control of Flocculation.” In *Yeast Genetics*, edited by J. F. T. Spencer, Dorothy M. Spencer, and A. R. W. Smith, 205–24. Springer Series in Molecular Biology. Springer New York. http://link.springer.com/chapter/10.1007/978-1-4612-5491-1_7.

Keane, Thomas M., Kim Wong, and David J. Adams. 2013. “RetroSeq: Transposable Element Discovery from next-Generation Sequencing Data.” *Bioinformatics* 29 (3): 389–90. doi:10.1093/bioinformatics/bts697.

Kendall, M. G. 1938. “A NEW MEASURE OF RANK CORRELATION.” *Biometrika* 30 (1–2): 81–93. doi:10.1093/biomet/30.1-2.81.

Kim, Hye Young, Sung Bae Lee, Hyen Sam Kang, Goo Taeg Oh, and TaeSoo Kim. 2014. “Two Distinct Domains of Flo8 Activator Mediates Its Role in Transcriptional Activation and the Physical Interaction with Mss11.” *Biochemical and Biophysical Research Communications* 449 (2): 202–7. doi:10.1016/j.bbrc.2014.04.161.

Kim, Yung-Jin, Leigh Francisco, Guang-Chao Chen, Edward Marcotte, and Clarence S. M. Chan. 1994. “Control of Cellular Morphogenesis by the Ip12/Bem2 GTPase-Activating Protein: Possible Role of Protein Phosphorylation.” *The Journal of Cell Biology* 127 (5): 1381–94.

Kojic, Erna M., and Rabih O. Darouiche. 2004. “Candida Infections of Medical Devices.” *Clinical Microbiology Reviews* 17 (2): 255–67. doi:10.1128/CMR.17.2.255-267.2004.

Koschwanez, John H, Kevin R Foster, and Andrew W Murray. 2013. “Improved Use of a Public Good Selects for the Evolution of Undifferentiated Multicellularity.” *eLife* 2 (April). doi:10.7554/eLife.00367.

Kucharczyk, R., R. Gromadka, A. Migdalski, P. P. Słonimski, and J. Rytka. 1999. “Disruption of Six Novel Yeast Genes Located on Chromosome II Reveals One Gene Essential for Vegetative Growth and Two Required for Sporulation and Conferring Hypersensitivity to Various Chemicals.” *Yeast* 15 (10B): 987–1000. doi:10.1002/(SICI)1097-0061(199907)15:10B<987::AID-YEA403>3.0.CO;2-5.

Kvitek, Daniel J., Jessica L. Will, and Audrey P. Gasch. 2008. “Variations in Stress Sensitivity and Genomic Expression in Diverse S. Cerevisiae Isolates.” *PLOS Genet* 4 (10): e1000223. doi:10.1371/journal.pgen.1000223.

Lander, E. S., and D. Botstein. 1989. “Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps.” *Genetics* 121 (1): 185–99.

Lander, E. S., and N. J. Schork. 1994. “Genetic Dissection of Complex Traits.” *Science* 265 (5181): 2037–48. doi:10.1126/science.8091226.

Layer, Ryan M., Colby Chiang, Aaron R. Quinlan, and Ira M. Hall. 2014. “LUMPY: A Probabilistic Framework for Structural Variant Discovery.” *Genome Biology* 15: R84. doi:10.1186/gb-2014-15-6-r84.

Lee, Hana N., Paul M. Magwene, and Rachel B. Brem. 2011. “Natural Variation in CDC28 Underlies Morphological Phenotypes in an Environmental Yeast Isolate.” *Genetics* 188 (3): 723–30. doi:10.1534/genetics.111.128819.

Legras, Jean-Luc, Didier Merdinoglu, Jean-Marie Cornuet, and Francis Karst. 2007. “Bread, Beer and Wine: Saccharomyces Cerevisiae Diversity Reflects Human History.” *Molecular Ecology* 16 (10): 2091–2102. doi:10.1111/j.1365-294X.2007.03266.x.

Leu, Jun-Yi, and Andrew W. Murray. 2006. “Experimental Evolution of Mating Discrimination in Budding Yeast.” *Current Biology* 16 (3): 280–86. doi:10.1016/j.cub.2005.12.028.

Li, Heng, and Richard Durbin. 2009. “Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform.” *Bioinformatics* 25 (14): 1754–60. doi:10.1093/bioinformatics/btp324.

Li, Xiaogang, Zhun Yan, and Jianping Xu. 2003. “Quantitative Variation of Biofilms among Strains in Natural Populations of Candida Albicans.” *Microbiology* 149 (2): 353–62. doi:10.1099/mic.0.25932-0.

Lipke, P. N., and C. Hull-Pillsbury. 1984. “Flocculation of Saccharomyces Cerevisiae tup1 Mutants.” *Journal of Bacteriology* 159 (2): 797–99.

Liti, Gianni, David B. H. Barton, and Edward J. Louis. 2006. “Sequence Diversity, Reproductive Isolation and Species Concepts in Saccharomyces.” *Genetics* 174 (2): 839–50. doi:10.1534/genetics.106.062166.

Liti, Gianni, David M. Carter, Alan M. Moses, Jonas Warringer, Leopold Parts, Stephen A. James, Robert P. Davey, et al. 2009. “Population Genomics of Domestic and Wild Yeasts.” *Nature* 458 (7236): 337–41. doi:10.1038/nature07743.

Liti, Gianni, and Edward J. Louis. 2012. “Advances in Quantitative Trait Analysis in Yeast.” *PLOS Genet* 8 (8): e1002912. doi:10.1371/journal.pgen.1002912.

Liti, Gianni, Antonella Peruffo, Steve A. James, Ian N. Roberts, and Edward J. Louis. 2005. "Inferences of Evolutionary Relationships from a Population Survey of LTR-Retrotransposons and Telomeric-Associated Sequences in the Saccharomyces Sensu Stricto Complex." *Yeast (Chichester, England)* 22 (3): 177–92. doi:10.1002/yea.1200.

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 (3): 967–78.

Lo, W. S., and A. M. Dranginis. 1996. "FLO11, a Yeast Gene Related to the STA Genes, Encodes a Novel Cell Surface Flocculin." *Journal of Bacteriology* 178 (24): 7144–51.

Lo, Wan-Sheng, and Anne M. Dranginis. 1998. "The Cell Surface Flocculin Flo11 Is Required for Pseudohyphae Formation and Invasion by Saccharomyces Cerevisiae." *Molecular Biology of the Cell* 9 (1): 161–71. doi:10.1091/mbc.9.1.161.

Lohse, Matthew B., Rebecca E. Zordan, Christopher W. Cain, and Alexander D. Johnson. 2010. "Distinct Class of DNA-Binding Domains Is Exemplified by a Master Regulator of Phenotypic Switching in Candida Albicans." *Proceedings of the National Academy of Sciences* 107 (32): 14105–10. doi:10.1073/pnas.1005911107.

Lorenz, Michael C., and Joseph Heitman. 1997. "Yeast Pseudohyphal Growth Is Regulated by GPA2, a G Protein α Homolog." *The EMBO Journal* 16 (23): 7008–18. doi:10.1093/emboj/16.23.7008.

Lou, Qiaojun, Liang Chen, Hanwei Mei, Haibin Wei, Fangjun Feng, Pei Wang, Hui Xia, Tiemei Li, and Lijun Luo. 2015. "Quantitative Trait Locus Mapping of Deep Rooting by Linkage and Association Analysis in Rice." *Journal of Experimental Botany* 66 (15): 4749–57. doi:10.1093/jxb/erv246.

Ludlow, Catherine L., Gareth A. Cromie, Cecilia Garmendia-Torres, Amy Sirr, Michelle Hays, Colburn Field, Eric W. Jeffery, Justin C. Fay, and Aimée M. Dudley. 2016. "Independent Origins of Yeast Associated with Coffee and Cacao Fermentation." *Current Biology* 26 (7): 965–71. doi:10.1016/j.cub.2016.02.012.

Mackay, Trudy F. C., Eric A. Stone, and Julien F. Ayroles. 2009. "The Genetics of Quantitative Traits: Challenges and Prospects." *Nature Reviews Genetics* 10 (8): 565–77. doi:10.1038/nrg2612.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20 (9): 1297–1303. doi:10.1101/gr.107524.110.

Merritt, Judith H., Daniel E. Kadouri, and George A. O'Toole. 2005. "Growing and Analyzing Static Biofilms." In *Current Protocols in Microbiology*. John Wiley & Sons, Inc. http://onlinelibrary.wiley.com/doi/10.1002/9780471729259.mc01b01s22/abstract.

Miki, B L, N H Poon, A P James, and V L Seligy. 1982. "Possible Mechanism for Flocculation Interactions Governed by Gene FLO1 in Saccharomyces Cerevisiae." *Journal of Bacteriology* 150 (2): 878–89.

Miller, Aaron W., Corrie Befort, Emily O. Kerr, and Maitreya J. Dunham. 2013. "Design and Use of Multiplexed Chemostat Arrays." *Journal of Visualized Experiments*, no. 72(February). doi:10.3791/50262.

Monod, J. 1950. "Technique, Theory and Applications of Continuous Culture." *Ann. Inst. Pasteur* 79 (4): 390–410.

Morillon, Antonin, Mathias Springer, and Pascale Lesage. 2000. "Activation of the Kss1 Invasive-Filamentous Growth Pathway Induces Ty1 Transcription and Retrotransposition inSaccharomyces Cerevisiae." *Molecular and Cellular Biology* 20 (15): 5766–76. doi:10.1128/MCB.20.15.5766-5776.2000.

Mösch, Hans-Ulrich, and Gerald R. Fink. 1997. "Dissection of Filamentous Growth by Transposon Mutagenesis in Saccharomyces Cerevisiae." *Genetics* 145 (3): 671–84.

Munson, R. J., and B. A. Bridges. 1964. "'Take-Over'—an Unusual Selection Process in Steady-State Cultures of Escherichia Coli." *Microbiology* 37 (3): 411–18. doi:10.1099/00221287-37-3-411.

Nilsson-Ehle, Herman. 1909. "Kreuzungsuntersuchungen an Hafer Und Weizen." http://echo.mpiwg-berlin.mpg.de/ECHOdocuView?url=/permanent/vlp/lit20073/index.meta.

Novick, Aaron, and Leo Szilard. 1950. "Description of the Chemostat." *Science* 112 (2920): 715–16. doi:10.1126/science.112.2920.715.

Orgogozo, Virginie. 2015. "Replaying the Tape of Life in the Twenty-First Century." *Interface Focus* 5 (6): 20150057. doi:10.1098/rsfs.2015.0057.

O'Toole, George A., and Roberto Kolter. 1998. "Initiation of Biofilm Formation in Pseudomonas Fluorescens WCS365 Proceeds via Multiple, Convergent Signalling Pathways: A Genetic Analysis." *Molecular Microbiology* 28 (3): 449–61. doi:10.1046/j.1365-2958.1998.00797.x.

O'Toole, George, Heidi B. Kaplan, and Roberto Kolter. 2000. "Biofilm Formation as Microbial Development." *Annual Review of Microbiology* 54 (1): 49–79. doi:10.1146/annurev.micro.54.1.49.

Oud, Bart, Victor Guadalupe-Medina, Jurgen F. Nijkamp, Dick de Ridder, Jack T. Pronk, Antonius J. A. van Maris, and Jean-Marc Daran. 2013. "Genome Duplication and Mutations in ACE2 Cause Multicellular, Fast-Sedimenting Phenotypes in Evolved Saccharomyces Cerevisiae." *Proceedings of the National Academy of Sciences* 110 (45): E4223–31. doi:10.1073/pnas.1305949110.

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 (3): 1005–23.

Palková, Zdena. 2004. "Multicellular Microorganisms: Laboratory versus Nature." *EMBO Reports* 5 (5): 470–76. doi:10.1038/sj.embor.7400145.

Papamichos-Chronakis, Manolis, R. Steven Conlan, Niki Gounalaki, Tjana Copf, and Dimitris Tzamarias. 2000. "Hrs1/Med3 Is a Cyc8-Tup1 Corepressor Target in the RNA Polymerase II Holoenzyme." *The Journal of Biological Chemistry* 275 (12): 8397-8403.

Pashkova, Natasha, Lokesh Gakhar, Stanley Winistorfer, Anna B. Sunshine, Matthew Rich, Maitreya J. Dunham, Liping Yu, and Robert Piper. 2013. "The Yeast Alix Homolog, Bro1, Functions as a Ubiquitin Receptor for Protein Sorting into Multivesicular Endosomes." *Developmental Cell* 25 (5): 520–33. doi:10.1016/j.devcel.2013.04.007.

Paterson, Andrew H., Eric S. Lander, John D. Hewitt, Susan Peterson, Stephen E. Lincoln, and Steven D. Tanksley. 1988. "Resolution of Quantitative Traits into Mendelian Factors by Using a Complete Linkage Map of Restriction Fragment Length Polymorphisms." *Nature* 335 (6192): 721–26. doi:10.1038/335721a0.

Poltak, Steffen R, and Vaughn S Cooper. 2011. "Ecological Succession in Long-Term Experimentally Evolved Biofilms Produces Synergistic Communities." *The ISME Journal* 5 (3): 369–78. doi:10.1038/ismej.2010.136.

Popescu, Andreia, and R. J. Doyle. 1996. “The Gram Stain after More than a Century.” *Biotechnic & Histochemistry* 71 (3): 145–51. doi:10.3109/10520299609117151.

Pringle, John R. 1991. “[52] Staining of Bud Scars and Other Cell Wall Chitin with Calcofluor.” In , edited by BT - Methods in Enzymology, 194:732–35. Guide to Yeast Genetics and Molecular Biology. Academic Press. http://www.sciencedirect.com/science/article/pii/007668799194055H.

Ramage, Gordon, Kacy Vande Walle, Brian L. Wickes, and José L. López-Ribot. 2001. “Standardized Method for In Vitro Antifungal Susceptibility Testing of Candida albicansBiofilms.” *Antimicrobial Agents and Chemotherapy* 45 (9): 2475–79. doi:10.1128/AAC.45.9.2475-2479.2001.

Renda, Brian A., Michael J. Hammerling, and Jeffrey E. Barrick. 2014. “Engineering Reduced Evolutionary Potential for Synthetic Biology” 10 (7): 1668–78. doi:10.1039/C3MB70606K.

Ratcliff, William C., Johnathon D. Fankhauser, David W. Rogers, Duncan Greig, and Michael Travisano. 2015. “Origins of Multicellular Evolvability in Snowflake Yeast.” *Nature Communications* 6 (January): 6102. doi:10.1038/ncomms7102.

Reynolds, Todd B. 2006. “The Opi1p Transcription Factor Affects Expression of FLO11, Mat Formation, and Invasive Growth in Saccharomyces Cerevisiae.” *Eukaryotic Cell* 5 (8): 1266–75. doi:10.1128/EC.00022-06.

Reynolds, Todd B., and Gerald R. Fink. 2001. “Bakers’ Yeast, a Model for Fungal Biofilm Formation.” *Science* 291 (5505): 878–81. doi:10.1126/science.291.5505.878.

Reynolds, Todd B., An Jansen, Xin Peng, and Gerald R. Fink. 2008. “Mat Formation in Saccharomyces Cerevisiae Requires Nutrient and pH Gradients.” *Eukaryotic Cell* 7 (1): 122–30. doi:10.1128/EC.00310-06.

Ribeiro, Suzana M., Mário R. Felício, Esther Vilas Boas, Sónia Gonçalves, Fabricio Costa, Ramar Perumal Samy, Nuno C. Santos, and Octávio L. Franco. 2016. “New Frontiers for Anti-Biofilm Drug Development.” *Pharmacology & Therapeutics*. Accessed February 22. doi:10.1016/j.pharmthera.2016.02.006.

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 & Development* 8 (24): 2974–85. doi:10.1101/gad.8.24.2974.

Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. “Integrative Genomics Viewer.” *Nature Biotechnology* 29 (1): 24–26. doi:10.1038/nbt.1754.

Roop, Jeremy I., and Rachel B. Brem. 2013. “Rare Variants in Hypermutable Genes Underlie Common Morphology and Growth Traits in Wild Saccharomyces Paradoxus.” *Genetics* 195 (2): 513–25. doi:10.1534/genetics.113.155341.

Rothstein, R. J., and F. Sherman. 1980. “Dependence on Mating Type for the Overproduction of Iso-2-Cytochrome c in the Yeast Mutant CYC7-H2.” *Genetics* 94 (4): 891–98.

Rowe, Steven M., Stacey Miller, and Eric J. Sorscher. 2005. “Cystic Fibrosis.” *New England Journal of Medicine* 352 (19): 1992–2001. doi:10.1056/NEJMra043184.

Ruusuvuori, Pekka, Jake Lin, Adrian C. Scott, Zhihao Tan, Saija Sorsa, Aleksi Kallio, Matti Nykter, Olli Yli-Harja, Ilya Shmulevich, and Aimée M. Dudley. 2014. “Quantitative Analysis of Colony Morphology in Yeast.” *BioTechniques* 56 (1): 18–27. doi:10.2144/000114123.

Ryan, Owen, Rebecca S. Shapiro, Christoph F. Kurat, David Mayhew, Anastasia Baryshnikova, Brian Chin, Zhen-Yuan Lin, et al. 2012. “Global Gene Deletion Analysis Exploring Yeast Filamentous Growth.” *Science* 337 (6100): 1353–56. doi:10.1126/science.1224339.

Saldanha, Alok J. 2004. “Java Treeview—extensible Visualization of Microarray Data.” *Bioinformatics* 20 (17): 3246–48. doi:10.1093/bioinformatics/bth349.

Salmon, Jean-Michel. 1997. “Enological Fermentation Kinetics of an Isogenic Ploidy Series Derived from an Industrial Saccharomyces Cerevisiae Strain.” *Journal of Fermentation and Bioengineering* 83 (3): 253–60. doi:10.1016/S0922-338X(97)80988-9.

Samayoa, Luis Fernando, Rosa Ana Malvar, Bode A Olukolu, James B Holland, and Ana Butrón. 2015. “Genome-Wide Association Study Reveals a Set of Genes Associated with Resistance to the Mediterranean Corn Borer (Sesamia Nonagrioides L.) in a Maize Diversity Panel.” *BMC Plant Biology* 15 (February). doi:10.1186/s12870-014-0403-3.

Sanchez, Monica Rose, Aaron W. Miller, Ivan Liachko, Anna B. Sunshine, Bryony Lynch, Mei Huang, Christopher G. DeSevo, et al. 2016. "Differential Paralog Divergence Modulates Evolutionary Outcomes in Yeast." *bioRxiv*, July, 63248. doi:10.1101/063248.

Sarode, Neha, Sarah E. Davis, Robert N. Tams, and Todd B. Reynolds. 2014. "The Wsc1p Cell Wall Signaling Protein Controls Biofilm (Mat) Formation Independently of Flo11p in Saccharomyces Cerevisiae." *G3: Genes|Genomes|Genetics* 4 (2): 199–207. doi:10.1534/g3.113.006361.

Sax, Karl. 1923. "The Association of Size Differences with Seed-Coat Pattern and Pigmentation in PHASEOLUS VULGARIS." *Genetics* 8 (6): 552–60.

Schacherer, Joseph, Joshua A. Shapiro, Douglas M. Ruderfer, and Leonid Kruglyak. 2009. "Comprehensive Polymorphism Survey Elucidates Population Structure of S. Cerevisiae." *Nature* 458 (7236): 342–45. doi:10.1038/nature07670.

Schlenke, Todd A., and David J. Begun. 2004. "Strong Selective Sweep Associated with a Transposon Insertion in Drosophila Simulans." *Proceedings of the National Academy of Sciences of the United States of America* 101 (6): 1626–31. doi:10.1073/pnas.0303793101.

Segrè, Ayellet V., Andrew W. Murray, and Jun-Yi Leu. 2006. "High-Resolution Mutation Mapping Reveals Parallel Experimental Evolution in Yeast." *PLOS Biol* 4 (8): e256. doi:10.1371/journal.pbio.0040256.

Selmecki, Anna M., Yosef E. Maruvka, Phillip A. Richmond, Marie Guillet, Noam Shoresh, Amber L. Sorenson, Subhajyoti De, et al. 2015. "Polyploidy Can Drive Rapid Adaptation in Yeast." *Nature* advance online publication (March). doi:10.1038/nature14187.

Servant, Géraldine, Carole Pennetier, and Pascale Lesage. 2008. "Remodeling Yeast Gene Transcription by Activating the Ty1 Long Terminal Repeat Retrotransposon under Severe Adenine Deficiency." *Molecular and Cellular Biology* 28 (17): 5543–54. doi:10.1128/MCB.00416-08.

Seshan, Venkatraman, and Adam Olshen. 2015. "DNAcopy: A Package for Analyzing DNA Copy Data." https://bioconductor.riken.jp/packages/3.2/bioc/vignettes/DNAcopy/inst/doc/DNAcopy.pdf.

Shen, Yu, Xiao Chen, Bingyin Peng, Liyuan Chen, Jin Hou, and Xiaoming Bao. 2012. "An Efficient Xylose-Fermenting Recombinant Saccharomyces Cerevisiae." *Applied Microbiology and Biotechnology* 96 (4): 1079–91. doi:10.1007/s00253-012-4418-0.

Sherman, Fred. 2002. "Getting Started with Yeast." In *Methods in Enzymology*, edited by Christine Guthrie and Gerald R. Fink, 350:3–41. Guide to Yeast Genetics and Molecular and Cell Biology - Part B. Academic Press. http://www.sciencedirect.com/science/article/pii/S007668790250954X.

Shoemaker, Daniel D., Deval A. Lashkari, Don Morris, Mike Mittmann, and Ronald W. Davis. 1996. "Quantitative Phenotypic Analysis of Yeast Deletion Mutants Using a Highly Parallel Molecular Bar–coding Strategy." *Nature Genetics* 14 (4): 450–56. doi:10.1038/ng1296-450.

Sinha, Himanshu, Lior David, Renata C. Pascon, Sandra Clauder-Münster, Sujatha Krishnakumar, Michelle Nguyen, Getao Shi, et al. 2008. "Sequential Elimination of Major-Effect Contributors Identifies Additional Quantitative Trait Loci Conditioning High-Temperature Growth in Yeast." *Genetics* 180 (3): 1661–70. doi:10.1534/genetics.108.092932.

Sivakumar, Ganapathy, Daniel R. Vail, Jianfeng Xu, David M. Burner, Jackson O. Lay, Xumeng Ge, and Pamela J. Weathers. 2010. "Bioethanol and Biodiesel: Alternative Liquid Fuels for Future Generations." *Engineering in Life Sciences* 10 (1): 8–18. doi:10.1002/elsc.200900061.

Skelly, Daniel A., Gennifer E. Merrihew, Michael Riffle, Caitlin F. Connelly, Emily O. Kerr, Marnie Johansson, Daniel Jaschob, et al. 2013. "Integrative Phenomics Reveals Insight into the Structure of Phenotypic Diversity in Budding Yeast." *Genome Research* 23 (9): 1496–1504. doi:10.1101/gr.155762.113.

Smukalla, Scott, Marina Caldara, Nathalie Pochet, Anne Beauvais, Stephanie Guadagnini, Chen Yan, Marcelo D. Vinces, et al. 2008. "FLO1 Is a Variable Green Beard Gene That Drives Biofilm-like Cooperation in Budding Yeast." *Cell* 135 (4): 726–37. doi:10.1016/j.cell.2008.09.037.

Sonderegger, Marco, and Uwe Sauer. 2003. "Evolutionary Engineering of Saccharomyces Cerevisiae for Anaerobic Growth on Xylose." *Applied and Environmental Microbiology* 69 (4): 1990–98. doi:10.1128/AEM.69.4.1990-1998.2003.

Stam, L. F., and C. C. Laurie. 1996. “Molecular Dissection of a Major Gene Effect on a Quantitative Trait: The Level of Alcohol Dehydrogenase Expression in Drosophila Melanogaster.” *Genetics* 144 (4): 1559–64.

Stark, Helene Cherrick, Donna Fugit, and Deborah Bernhardt Mowshowitz. 1980. “Pleiotropic Properties of a Yeast Mutant Insensitive to Catabolite Repression.” *Genetics* 94 (4): 921–28.

Steensels, Jan, and Kevin J. Verstrepen. 2014. “Taming Wild Yeast: Potential of Conventional and Nonconventional Yeasts in Industrial Fermentations.” *Annual Review of Microbiology* 68 (1): 61–80. doi:10.1146/annurev-micro-091213-113025.

Steinmetz, Lars M., Himanshu Sinha, Dan R. Richards, Jamie I. Spiegelman, Peter J. Oefner, John H. McCusker, and Ronald W. Davis. 2002. “Dissecting the Architecture of a Quantitative Trait Locus in Yeast.” *Nature* 416 (6878): 326–30. doi:10.1038/416326a.

Stemmer, W P. 1994. “DNA Shuffling by Random Fragmentation and Reassembly: In Vitro Recombination for Molecular Evolution.” *Proceedings of the National Academy of Sciences of the United States of America* 91 (22): 10747–51.

Storey, John D. 2002. “A Direct Approach to False Discovery Rates.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (3): 479–98. doi:10.1111/1467-9868.00346.

Šťovíček, Vratislav, Libuše Váchová, Martin Kuthan, and Zdena Palková. 2010. “General Factors Important for the Formation of Structured Biofilm-like Yeast Colonies.” *Fungal Genetics and Biology* 47 (12): 1012–22. doi:10.1016/j.fgb.2010.08.005.

Stratford, Malcolm. 1989. “Yeast Flocculation: Calcium Specificity.” *Yeast* 5 (6): 487–96. doi:10.1002/yea.320050608.

Stratford, Malcolm. 1992. “Lectin-Mediated Aggregation of Yeasts — Yeast Flocculation.” *Biotechnology and Genetic Engineering Reviews* 10 (1): 283–342. doi:10.1080/02648725.1992.10647891.

Sudarsanam, Priya, and Barak A. Cohen. 2014. “Single Nucleotide Variants in Transcription Factors Associate More Tightly with Phenotype than with Gene Expression.” *PLoS Genet* 10 (5): e1004325. doi:10.1371/journal.pgen.1004325.

Sunshine, Anna B., Celia Payen, Giang T. Ong, Ivan Liachko, Kean Ming Tan, and Maitreya J. Dunham. 2015. "The Fitness Consequences of Aneuploidy Are Driven by Condition-Dependent Gene Effects." *PLoS Biology* 13 (5): e1002155. doi:10.1371/journal.pbio.1002155.

Taylor, Matthew B., and Ian M. Ehrenreich. 2014. "Genetic Interactions Involving Five or More Genes Contribute to a Complex Trait in Yeast." *PLoS Genet* 10 (5): e1004324. doi:10.1371/journal.pgen.1004324.

Taylor, Matthew B., Joann Phan, Jonathan T. Lee, Madelyn McCadden, and Ian M. Ehrenreich. 2016. "Diverse Genetic Architectures Lead to the Same Cryptic Phenotype in a Yeast Cross." *Nature Communications* 7 (June): 11669. doi:10.1038/ncomms11669.

Tenaillon, Olivier, Jeffrey E. Barrick, Noah Ribeck, Daniel E. Deatherage, Jeffrey L. Blanchard, Aurko Dasgupta, Gabriel C. Wu, et al. 2016. "Tempo and Mode of Genome Evolution in a 50,000-Generation Experiment." *Nature* 536 (7615): 165–70. doi:10.1038/nature18959.

Teunissen, A. W., J. A. van den Berg, and H. Y. Steensma. 1995. "Transcriptional Regulation of Flocculation Genes in Saccharomyces Cerevisiae." *Yeast (Chichester, England)* 11 (5): 435–46. doi:10.1002/yea.320110506.

Thomas, D. N., and G. S. Dieckmann. 2002. "Antarctic Sea Ice--a Habitat for Extremophiles." *Science* 295 (5555): 641–44. doi:10.1126/science.1063391.

Todeschini, Anne-Laure, Antonin Morillon, Mathias Springer, and Pascale Lesage. 2005. "Severe Adenine Starvation Activates Ty1 Transcription and Retrotransposition in Saccharomyces Cerevisiae." *Molecular and Cellular Biology* 25 (17): 7459–72. doi:10.1128/MCB.25.17.7459-7472.2005.

Topiwala, H. H., and G. Hamer. 1971. "Effect of Wall Growth in Steady-State Continuous Cultures." *Biotechnology and Bioengineering* 13 (6): 919–22. doi:10.1002/bit.260130614.

Toprak, Erdal, Adrian Veres, Sadik Yildiz, Juan M. Pedraza, Remy Chait, Johan Paulsson, and Roy Kishony. 2013. "Building a Morbidostat: An Automated Continuous-Culture Device for Studying Bacterial Drug Resistance under Dynamically Sustained Drug Inhibition." *Nature Protocols* 8 (3): 555–67. doi:10.1038/nprot.2013.021.

Turk, Martina, Zrinka Abramović, Ana Plemenitaš, and Nina Gunde-Cimerman. 2007. "Salt Stress and Plasma-Membrane Fluidity in Selected Extremophilic Yeasts and Yeast-like Fungi." *FEMS Yeast Research* 7 (4): 550–57. doi:10.1111/j.1567-1364.2007.00209.x.

Váchová, Libuše, Vratislav Šťovíček, Otakar Hlaváček, Oleksandr Chernyavskiy, Luděk Štěpánek, Lucie Kubínová, and Zdena Palková. 2011. "Flo11p, Drug Efflux Pumps, and the Extracellular Matrix Cooperate to Form Biofilm Yeast Colonies." *The Journal of Cell Biology* 194 (5): 679–87. doi:10.1083/jcb.201103129.

Verstrepen, K. J., G. Derdelinckx, H. Verachtert, and F. R. Delvaux. 2003. "Yeast Flocculation: What Brewers Should Know." *Applied Microbiology and Biotechnology* 61 (3): 197–205. doi:10.1007/s00253-002-1200-8.

Verstrepen, Kevin J., An Jansen, Fran Lewitter, and Gerald R. Fink. 2005. "Intragenic Tandem Repeats Generate Functional Variability." *Nature Genetics* 37 (9): 986–90. doi:10.1038/ng1618.

Verstrepen, Kevin J., and Frans M. Klis. 2006. "Flocculation, Adhesion and Biofilm Formation in Yeasts." *Molecular Microbiology* 60 (1): 5–15. doi:10.1111/j.1365-2958.2006.05072.x.

Verstrepen, Kevin J., Todd B. Reynolds, and Gerald R. Fink. 2004. "Origins of Variation in the Fungal Cell Surface." *Nature Reviews Microbiology* 2 (7): 533–40. doi:10.1038/nrmicro927.

Voordeckers, Karin, Dries De Maeyer, Elisa van der Zande, Marcelo D. Vinces, Wim Meert, Lore Cloots, Owen Ryan, Kathleen Marchal, and Kevin J. Verstrepen. 2012. "Identification of a Complex Genetic Network Underlying Saccharomyces Cerevisiae Colony Morphology." *Molecular Microbiology* 86 (1): 225–39. doi:10.1111/j.1365-2958.2012.08192.x.

Voordeckers, Karin, and Kevin J Verstrepen. 2015. "Experimental Evolution of the Model Eukaryote Saccharomyces Cerevisiae Yields Insight into the Molecular Mechanisms Underlying Adaptation." *Current Opinion in Microbiology*, Growth and development: eukaryotes and prokaryotes, 28 (December): 1–9. doi:10.1016/j.mib.2015.06.018.

Vopálenská, Irena, Marta Hůlková, Blanka Janderová, and Zdena Palková. 2005. "The Morphology of Saccharomyces Cerevisiae Colonies Is Affected by Cell Adhesion and the Budding Pattern." *Research in Microbiology* 156 (9): 921–31. doi:10.1016/j.resmic.2005.05.012.

Vopálenská, Irena, Vratislav Šťovíček, Blanka Janderová, Libuše Váchová, and Zdena Palková. 2010. "Role of Distinct Dimorphic Transitions in Territory Colonizing and Formation of Yeast Colony Architecture." *Environmental Microbiology* 12 (1): 264–77. doi:10.1111/j.1462-2920.2009.02067.x.

Voth, Warren P., Aileen E. Olsen, Mohammed Sbia, Karen H. Freedman, and David J. Stillman. 2005. "ACE2, CBK1, and BUD4 in Budding and Cell Separation." *Eukaryotic Cell* 4 (6): 1018–28. doi:10.1128/EC.4.6.1018-1028.2005.

Wang, Yiting, Joe A. Wrennall, Zhiwei Cai, Hongyu Li, and David N. Sheppard. 2014. "Understanding How Cystic Fibrosis Mutations Disrupt CFTR Function: From Single Molecules to Animal Models." *The International Journal of Biochemistry & Cell Biology*, Cystic Fibrosis: From o-mics to cell biology, physiology, and therapeutic advances, 52 (July): 47–57. doi:10.1016/j.biocel.2014.04.001.

Warringer, Jonas, Enikö Zörgö, Francisco A. Cubillos, Amin Zia, Arne Gjuvsland, Jared T. Simpson, Annabelle Forsmark, et al. 2011. "Trait Variation in Yeast Is Defined by Population History." *PLOS Genet* 7 (6): e1002111. doi:10.1371/journal.pgen.1002111.

Wilke, C. M., and J. Adams. 1992. "Fitness Effects of Ty Transposition in Saccharomyces Cerevisiae." *Genetics* 131 (1): 31–42.

Wilkening, Stefan, Gen Lin, Emilie S. Fritsch, Manu M. Tekkedil, Simon Anders, Raquel Kuehn, Michelle Nguyen, et al. 2014. "An Evaluation of High-Throughput Approaches to QTL Mapping in Saccharomyces Cerevisiae." *Genetics* 196 (3): 853–65. doi:10.1534/genetics.113.160291.

Williams, F. E., and R. J. Trumbly. 1990. "Characterization of TUP1, a Mediator of Glucose Repression in Saccharomyces Cerevisiae." *Molecular and Cellular Biology* 10 (12): 6500–6511. doi:10.1128/MCB.10.12.6500.

Winkler, James D., and Katy C. Kao. 2014. "Recent Advances in the Evolutionary Engineering of Industrial Biocatalysts." *Genomics*, Experimental evolution and the use of genomics, 104 (6, Part A): 406–11. doi:10.1016/j.ygeno.2014.09.006.

Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, et al. 1999. “Functional Characterization of the S. Cerevisiae Genome by Gene Deletion and Parallel Analysis.” *Science (New York, N.Y.)* 285 (5429): 901–6.

Wohlbach, Dana J., Nikolay Rovinskiy, Jeffrey A. Lewis, Maria Sardi, Wendy S. Schackwitz, Joel A. Martin, Shweta Deshpande, et al. 2014. “Comparative Genomics of Saccharomyces Cerevisiae Natural Isolates for Bioenergy Production.” *Genome Biology and Evolution* 6 (9): 2557–66. doi:10.1093/gbe/evu199.

Wood, Andrew R., Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H. Pers, Stefan Gustafsson, Audrey Y. Chu, et al. 2014. “Defining the Role of Common Variation in the Genomic and Biological Architecture of Adult Human Height.” *Nature Genetics* 46 (11): 1173–86. doi:10.1038/ng.3097.

Wright, Sewell. 1968. “Evolution and the Genetics of Populations. Vol. 1. Genetic and Biométrie Foundations.” *Evolution and the Genetics of Populations. Vol. 1. Genetic and Biometrie Foundations.*

Xu, H, and J D Boeke. 1991. “Inhibition of Ty1 Transposition by Mating Pheromones in Saccharomyces Cerevisiae.” *Molecular and Cellular Biology* 11 (5): 2736–43.

Yokobayashi, Yohei, Ron Weiss, and Frances H. Arnold. 2002. “Directed Evolution of a Genetic Circuit.” *Proceedings of the National Academy of Sciences* 99 (26): 16587–91. doi:10.1073/pnas.252535999.

Yvert, Gaël, Rachel B. Brem, Jacqueline Whittle, Joshua M. Akey, Eric Foss, Erin N. Smith, Rachel Mackelprang, and Leonid Kruglyak. 2003. “Trans-Acting Regulatory Variation in Saccharomyces Cerevisiae and the Role of Transcription Factors.” *Nature Genetics* 35 (1): 57–64. doi:10.1038/ng1222.

Zara, Giacomo, Severino Zara, Claudia Pinna, Salvatore Marceddu, and Marilena Budroni. 2009. “FLO11 Gene Length and Transcriptional Level Affect Biofilm-Forming Ability of Wild Flor Strains of Saccharomyces Cerevisiae.” *Microbiology* 155 (12): 3838–46. doi:10.1099/mic.0.028738-0.

Zhang, Shicheng, Tianlong Zhang, Minghui Yan, Jianping Ding, and Jiangye Chen. 2014. “Crystal Structure of the WOPR-DNA Complex and Implications for Wor1 Function in White-Opaque Switching of Candida Albicans.” *Cell Research* 24 (9): 1108–20. doi:10.1038/cr.2014.102.

Zhang, Zhizhou, Ushasri Varanasi, and Robert J. Trumbly. 2002. “Functional Dissection of the Global Repressor Tup1 in Yeast: Dominant Role of the C-Terminal Repression Domain.” *Genetics* 161 (3): 957–69.

Zhao, Huimin, and Frances H. Arnold. 1997. “Optimization of DNA Shuffling for High Fidelity Recombination.” *Nucleic Acids Research* 25 (6): 1307–8. doi:10.1093/nar/25.6.1307.

Zhou, Hang, Jing-sheng Cheng, Benjamin L. Wang, Gerald R. Fink, and Gregory Stephanopoulos. 2012. “Xylose Isomerase Overexpression along with Engineering of the Pentose Phosphate Pathway and Evolutionary Engineering Enable Rapid Xylose Utilization and Ethanol Production by Saccharomyces Cerevisiae.” *Metabolic Engineering* 14 (6): 611–22. doi:10.1016/j.ymben.2012.07.011.

Zhou, Junli, Xiangfeng Wang, Yuling Jiao, Yonghua Qin, Xigang Liu, Kun He, Chen Chen, et al. 2007. “Global Genome Expression Analysis of Rice in Response to Drought and High-Salinity Stresses in Shoot, Flag Leaf, and Panicle.” *Plant Molecular Biology* 63 (5): 591–608. doi:10.1007/s11103-006-9111-1.

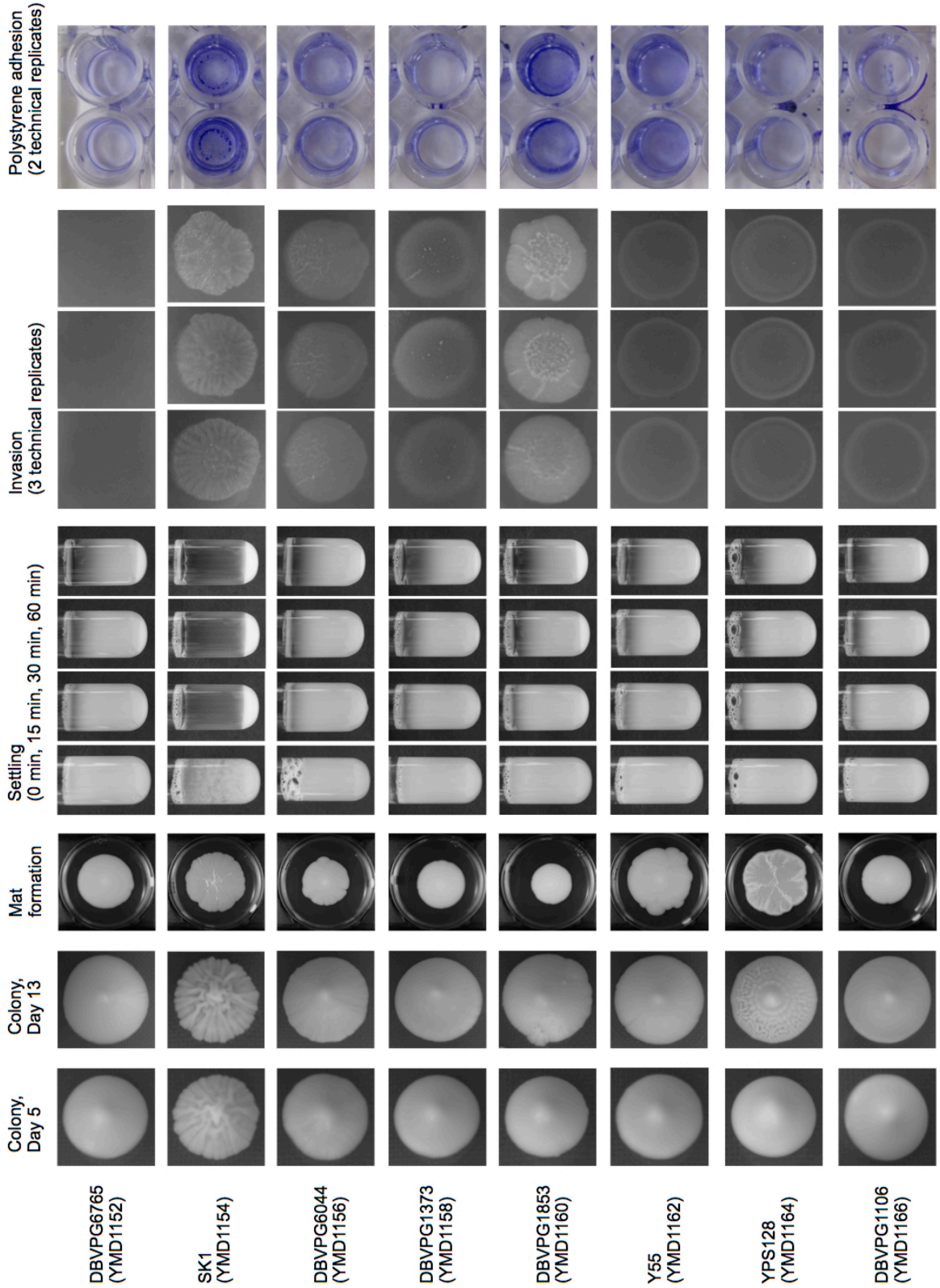
Zordan, Rebecca E., David J. Galgoczy, and Alexander D. Johnson. 2006. “Epigenetic Properties of White–opaque Switching in Candida Albicans Are Based on a Self-Sustaining Transcriptional Feedback Loop.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (34): 12807–12. doi:10.1073/pnas.0605138103.

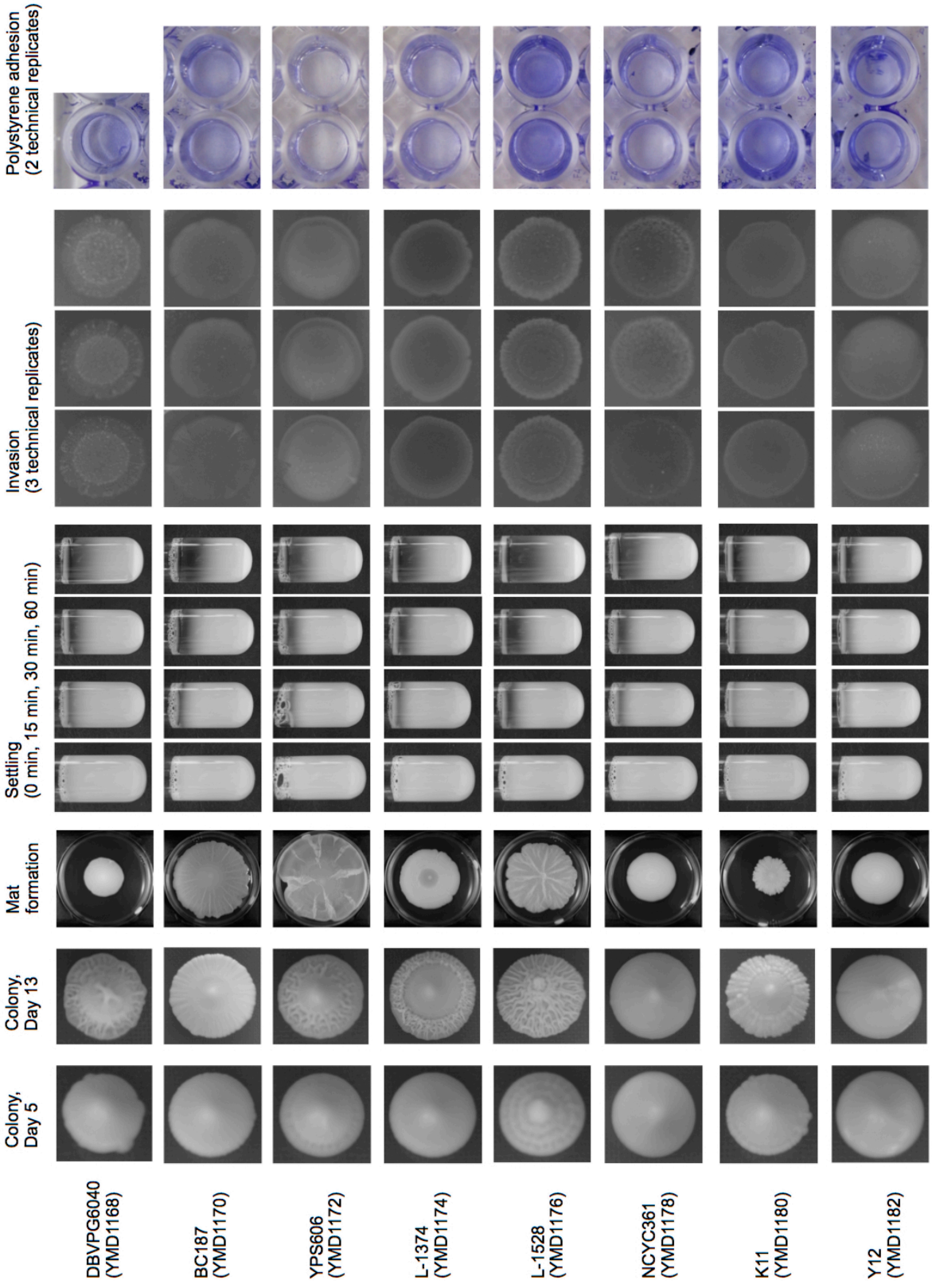
Zörgö, Enikö, Karolina Chwialkowska, Arne B. Gjuvsland, Elena Garré, Per Sunnerhagen, Gianni Liti, Anders Blomberg, Stig W. Omholt, and Jonas Warringer. 2013. “Ancient Evolutionary Trade-Offs between Yeast Ploidy States.” *PLOS Genet* 9 (3): e1003388. doi:10.1371/journal.pgen.1003388.

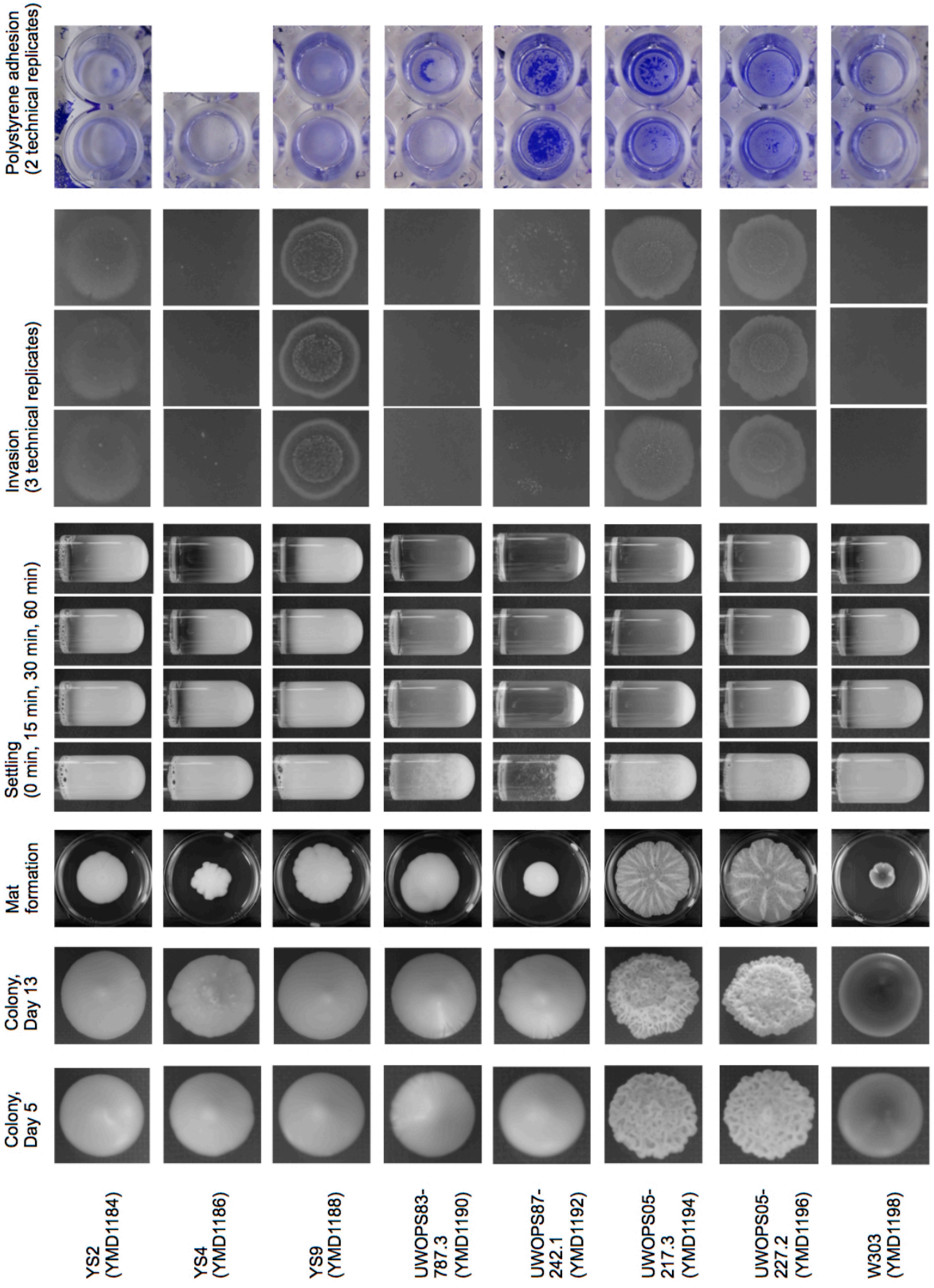
Zuk, Or, Eliana Hechter, Shamil R. Sunyaev, and Eric S. Lander. 2012. “The mystery of missing heritability: Genetic interactions create phantom heritability.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (4): 1193-98.

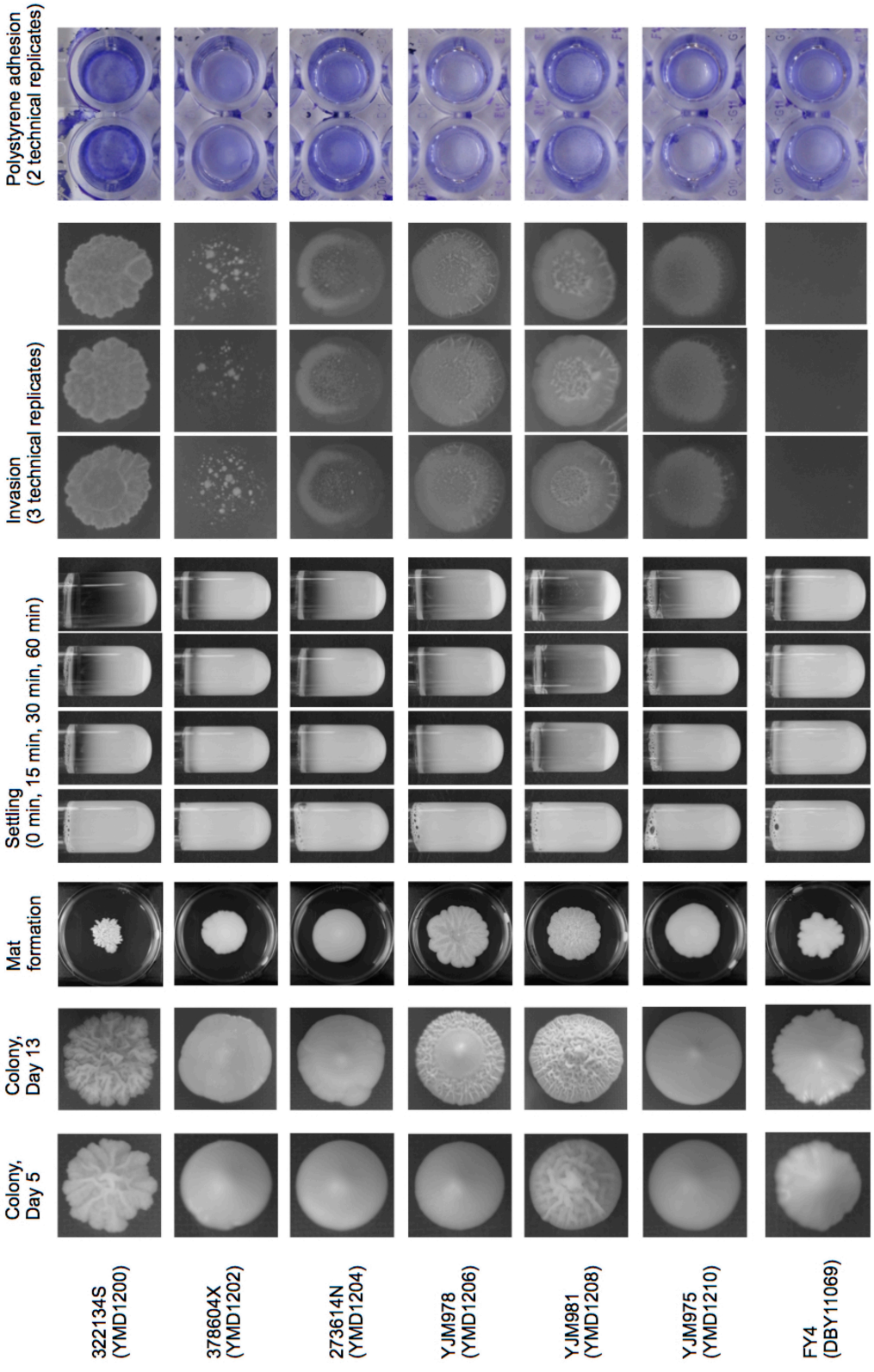
APPENDIX A

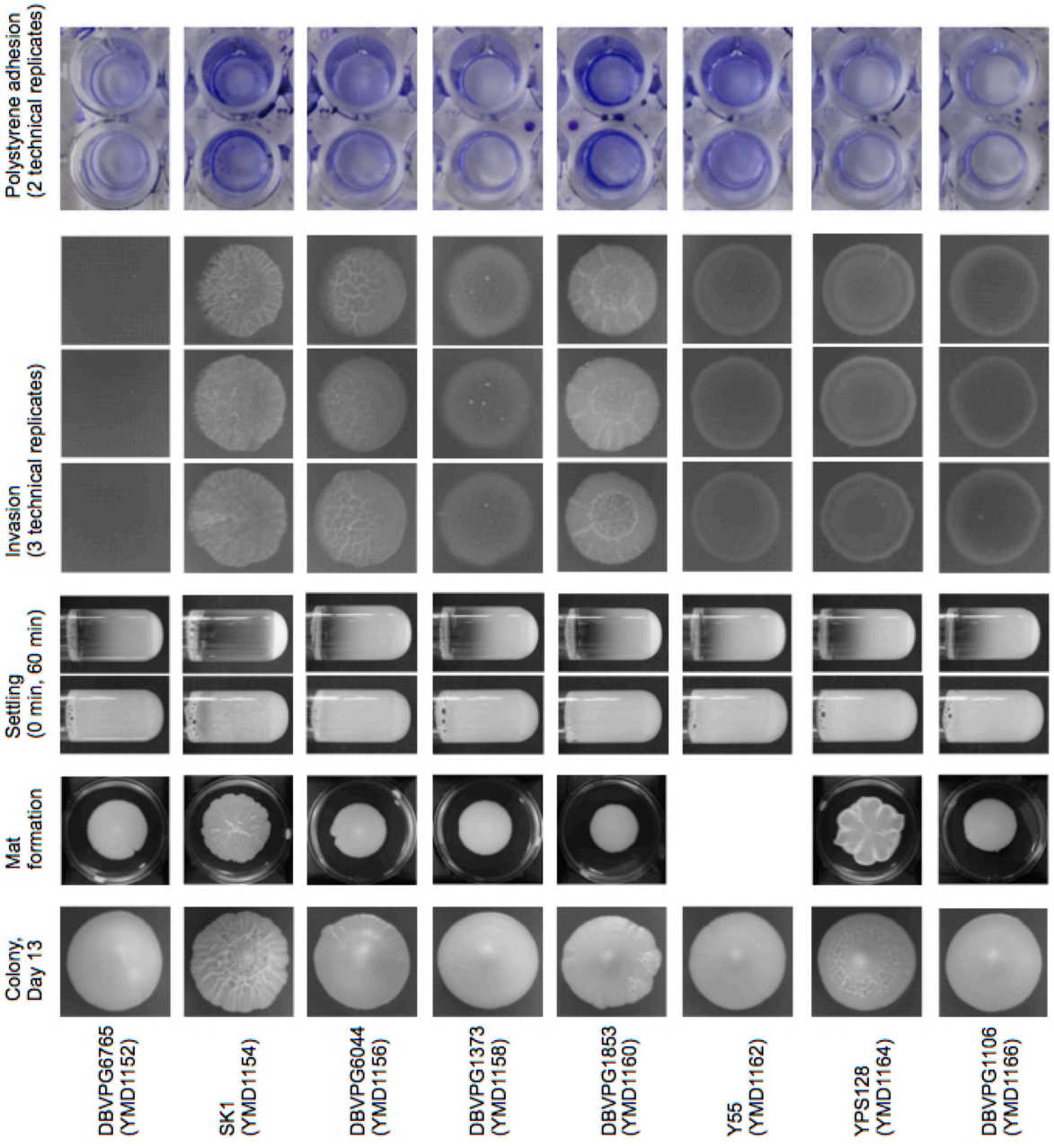
Figure A.1. Complete haploid phenotypic panel

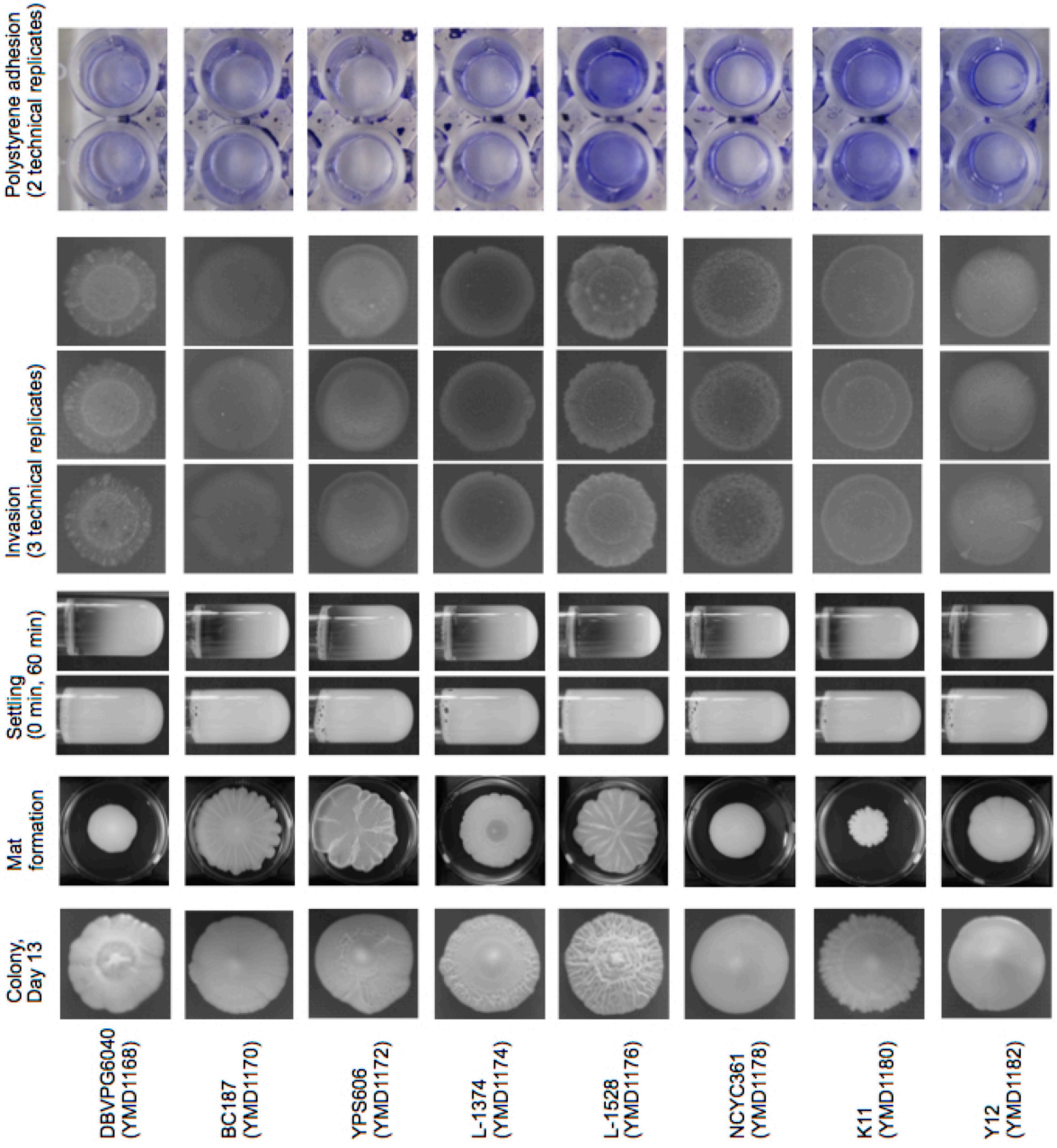


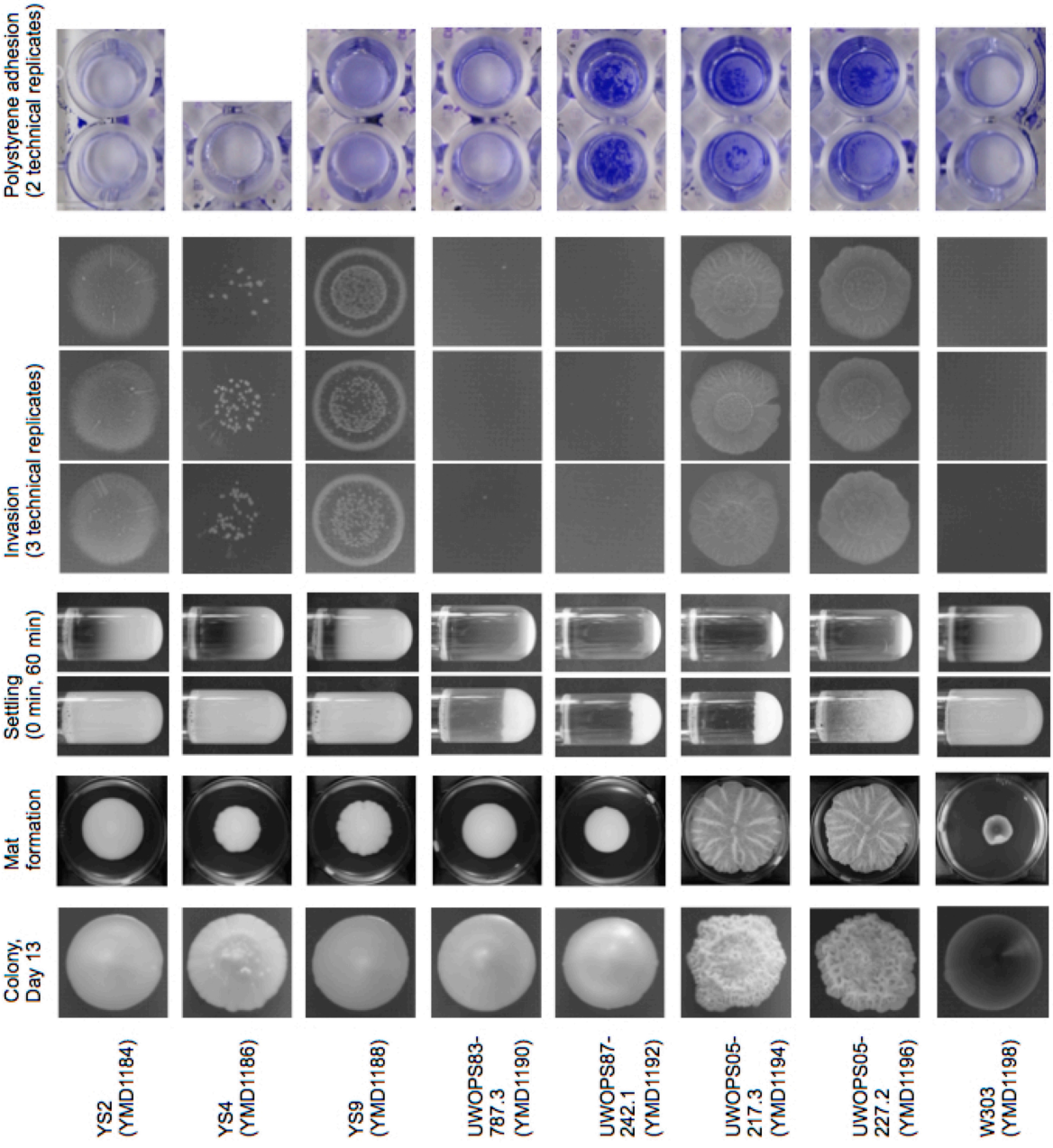

Colony, Day 5
Colony, Day 13
Mat formation
Settling (0 min, 15 min, 30 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical replicates)
DBVPG6040 (YMD1168)
BC187 (YMD1170)
YPS606 (YMD1172)
L-1374 (YMD1174)
L-1528 (YMD1176)
NCYC361 (YMD1178)
K11 (YMD1180)
Y12 (YMD1182)

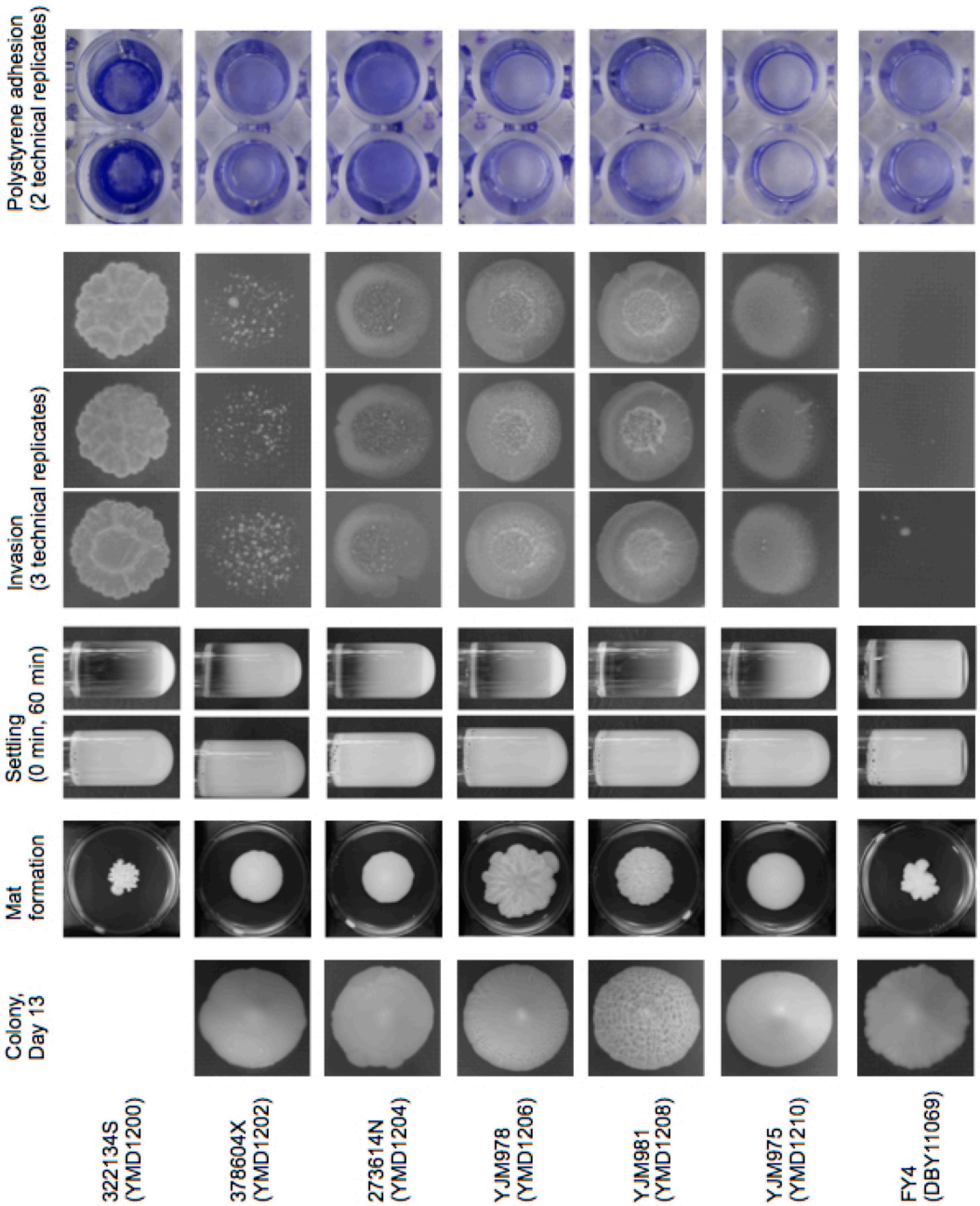

Colony, Day 5
Colony, Day 13
Mat formation
Settling (0 min, 15 min, 30 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical replicates)
YS2 (YMD1184)
YS4 (YMD1186)
YS9 (YMD1188)
UWOPS83-787.3 (YMD1190)
UWOPS87-242.1 (YMD1192)
UWOPS05-217.3 (YMD1194)
UWOPS05-227.2 (YMD1196)
W303 (YMD1198)


Colony, Day 5
Colony, Day 13
Mat formation
Settling (0 min, 15 min, 30 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical replicates)
322134S (YMD1200)
378604X (YMD1202)
273614N (YMD1204)
YJM978 (YMD1206)
YJM981 (YMD1208)
YJM975 (YMD1210)
FY4 (DBY11069)


Colony, Day 13
Mat formation
Settling (0 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical replicates)
DBVPG6765 (YMD1152)
SK1 (YMD1154)
DBVPG6044 (YMD1156)
DBVPG1373 (YMD1158)
DBVPG1853 (YMD1160)
Y55 (YMD1162)
YPS128 (YMD1164)
DBVPG1106 (YMD1166)

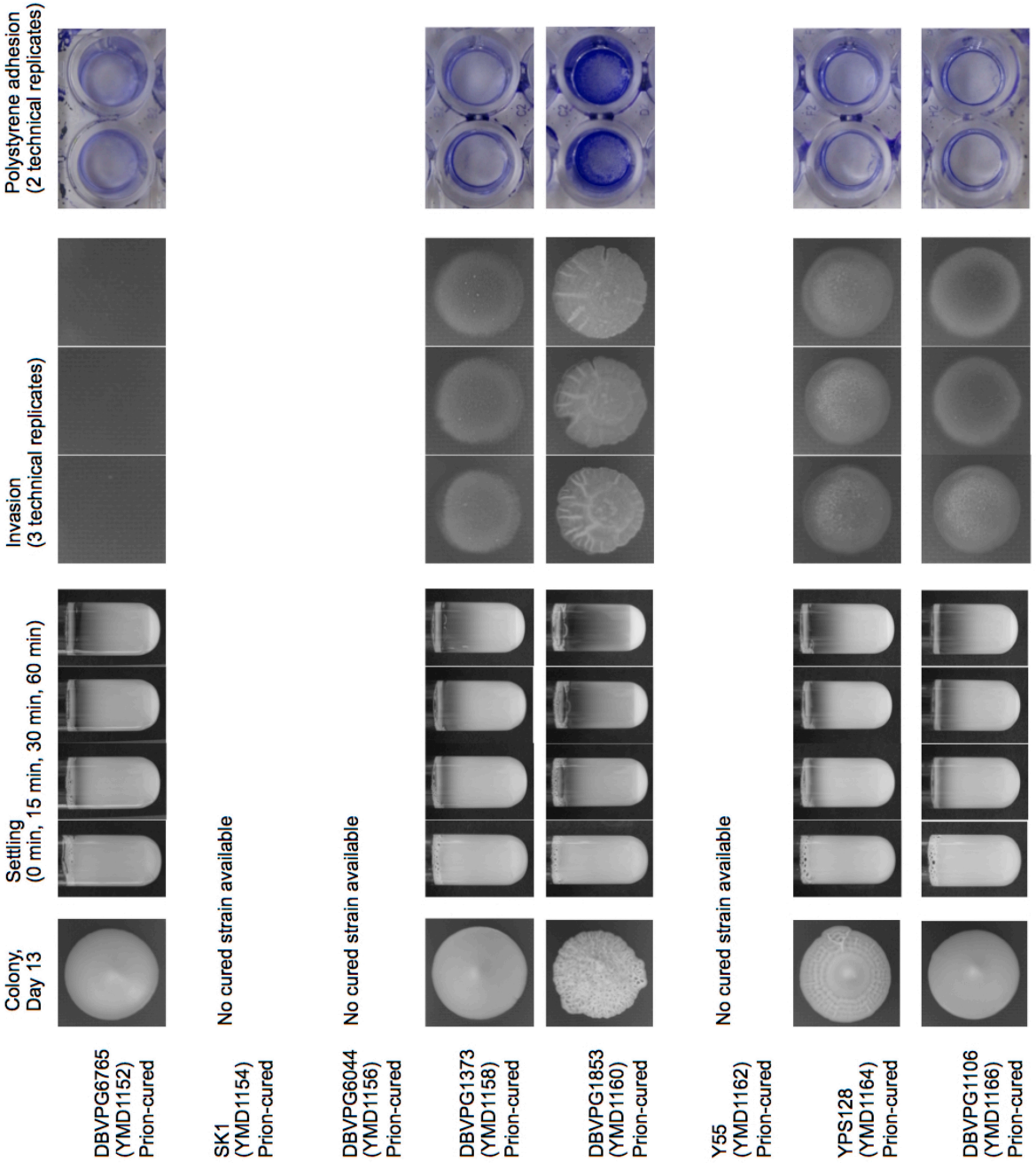

Colony, Day 13
Mat formation
Settling (0 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical replicates)
DBVPG6040 (YMD1168)
BC187 (YMD1170)
YPS606 (YMD1172)
L-1374 (YMD1174)
L-1528 (YMD1176)
NCYC361 (YMD1178)
K11 (YMD1180)
Y12 (YMD1182)

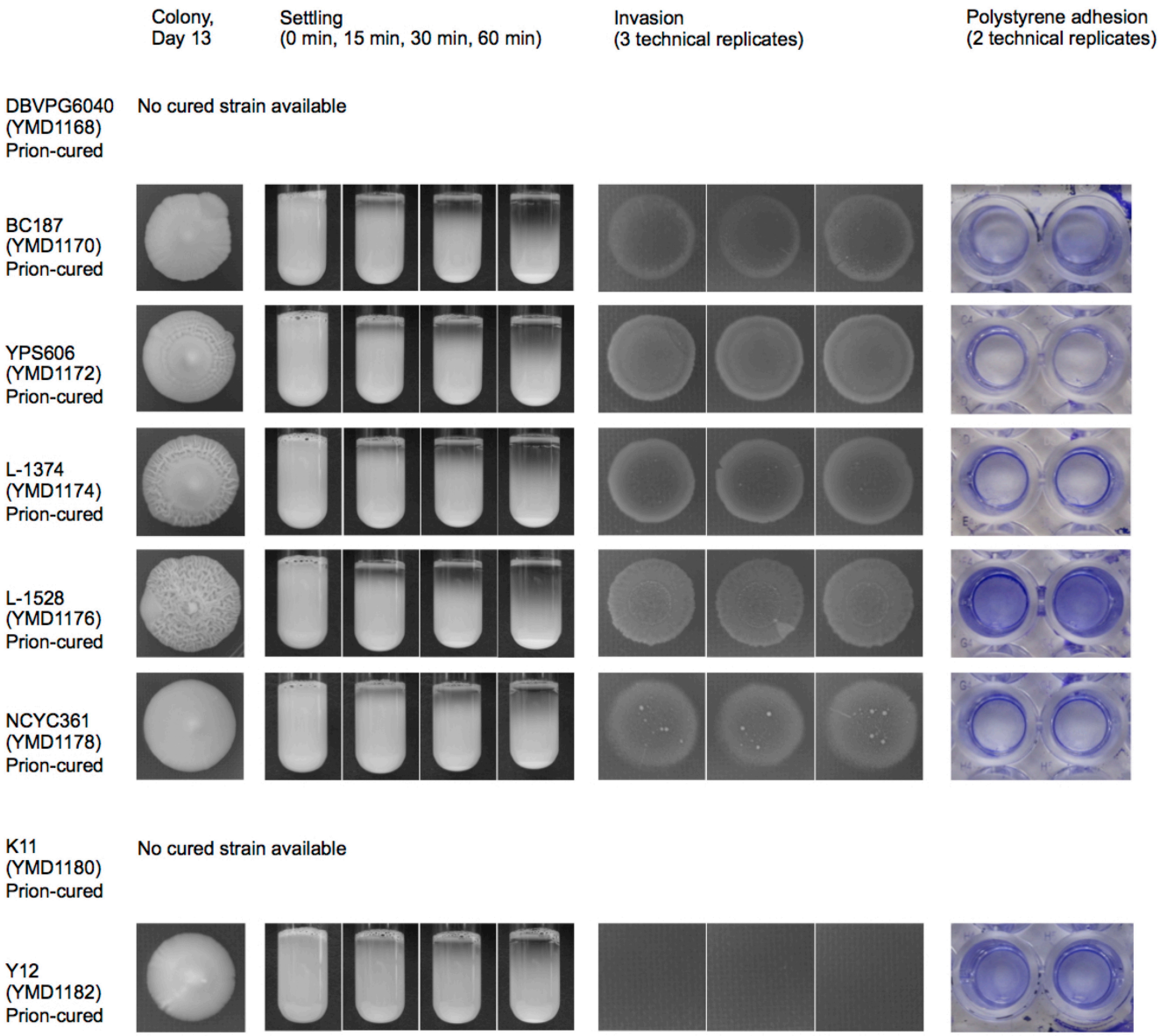

Polystyrene adhesion
(2 technical replicates)
Invasion
(3 technical replicates)
Settling
(0 min, 60 min)
Mat
formation
Colony,
Day 13
YS2
(YMD1184)
YS4
(YMD1186)
YS9
(YMD1188)
UWOPS83-
787.3
(YMD1190)
UWOPS87-
242.1
(YMD1192)
UWOPS05-
217.3
(YMD1194)
UWOPS05-
227.2
(YMD1196)
W303
(YMD1198)

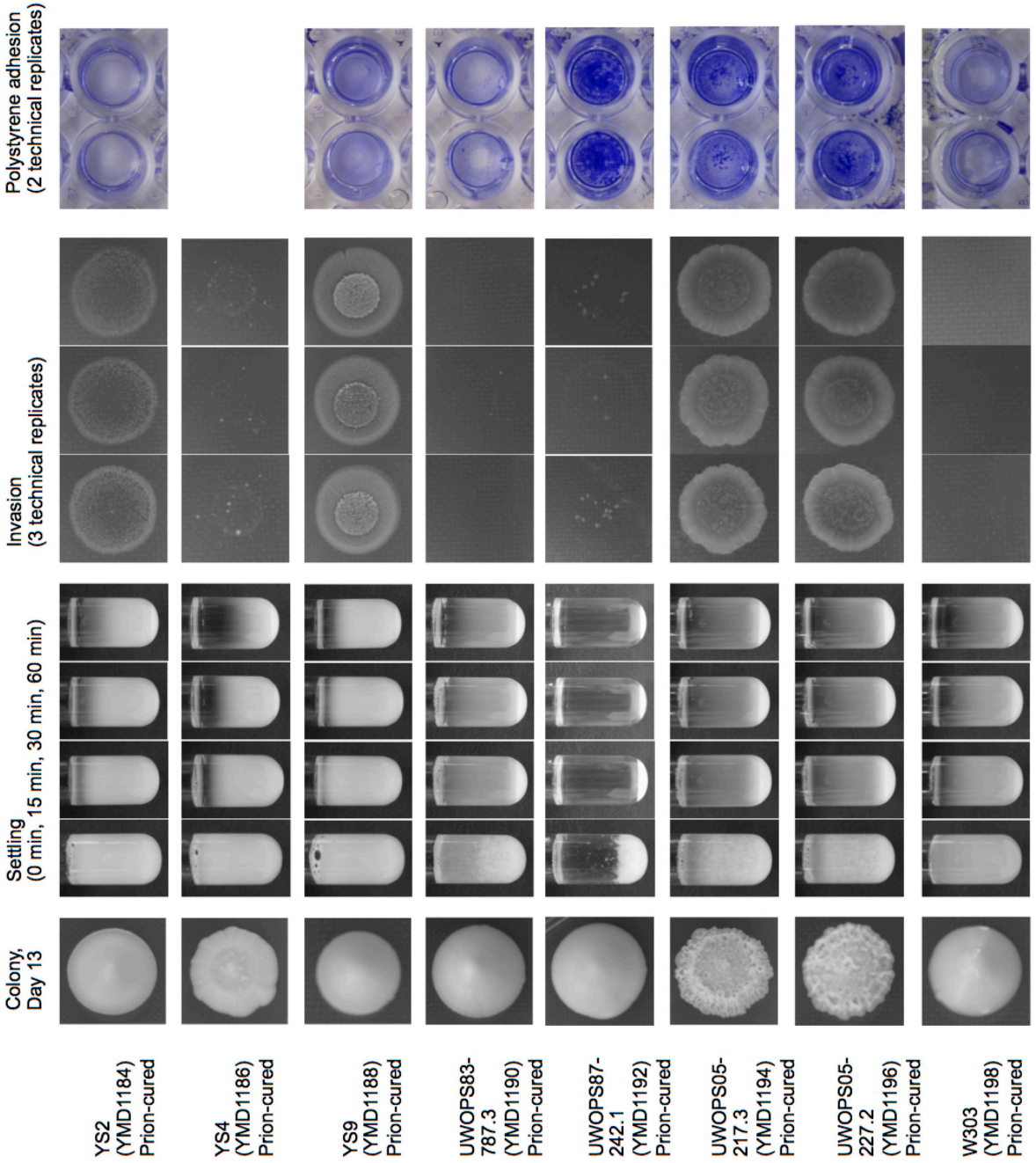


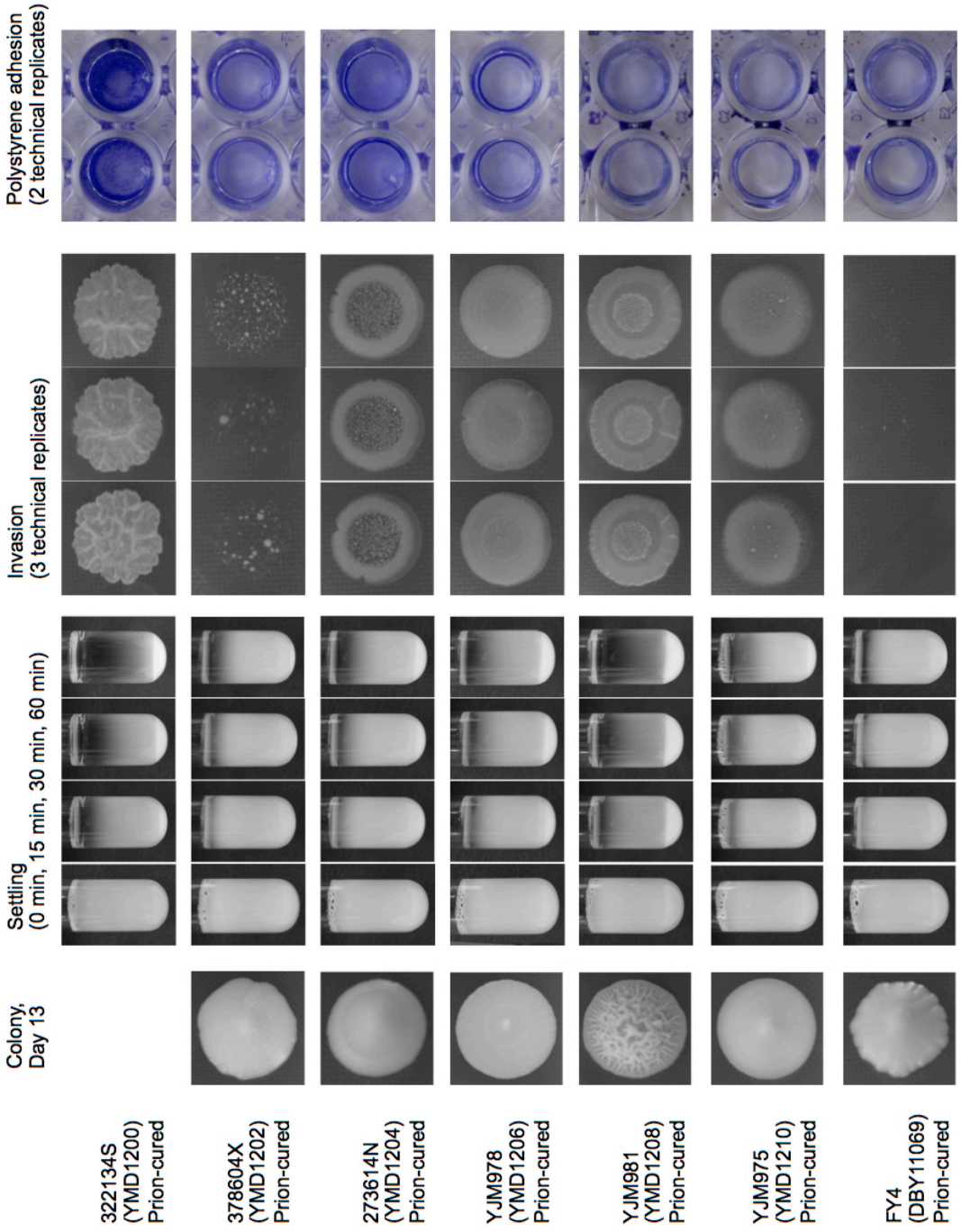
Full phenotypic panel for all 31 haploid strains used in study, in two biological replicates. Strains are listed with their formal name and origin and shown across five different phenotypes. Complex mat formation images include the plate for scale. Three technical replicates are shown for the invasion assay, photographed after 24h growth following washing on day 5. Two technical replicates are shown for the polystyrene adhesion assay. Pictured biofilms are fixed and stained with a 1% w/v crystal violet solution.

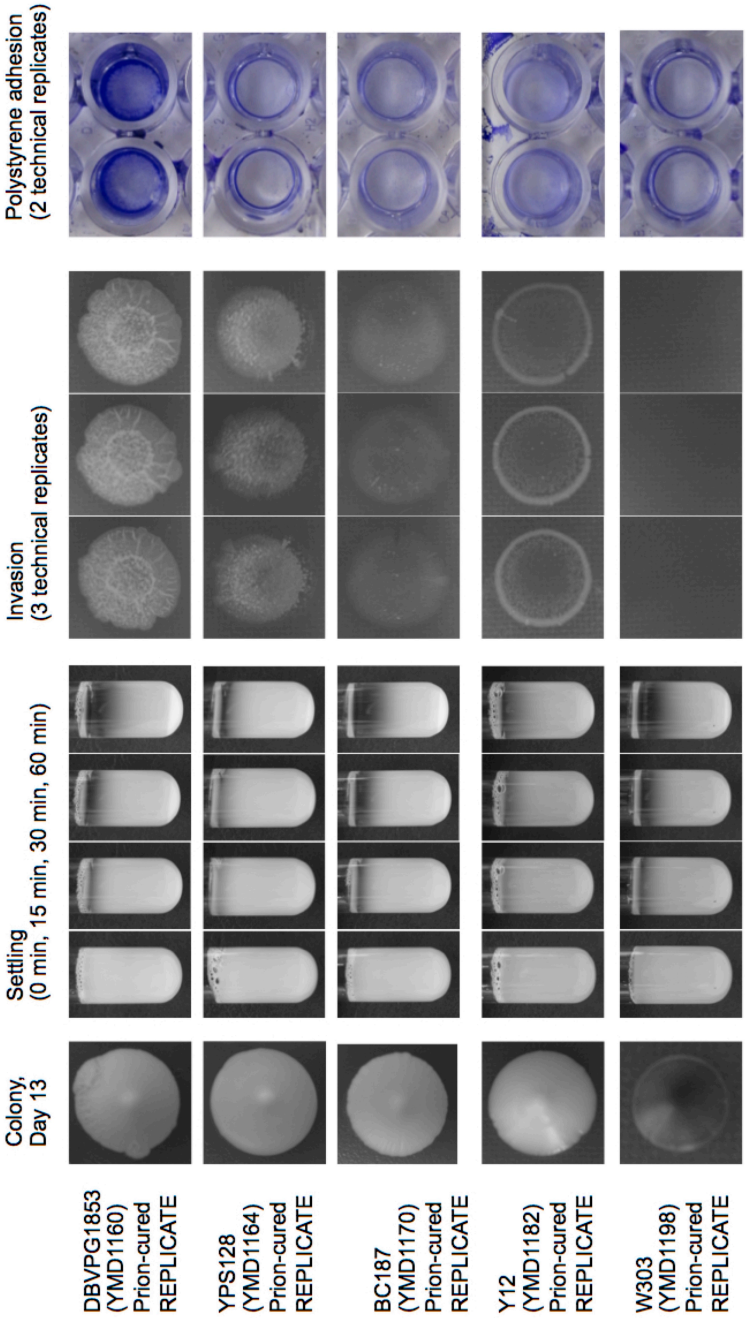
Figure A.2. Complete prion-cured haploid phenotypic panel




Colony, Day 13
Settling (0 min, 15 min, 30 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical replicates)
DBVPG6040 (YMD1168) Prion-cured
No cured strain available
BC187 (YMD1170) Prion-cured
YPS606 (YMD1172) Prion-cured
L-1374 (YMD1174) Prion-cured
L-1528 (YMD1176) Prion-cured
NCYC361 (YMD1178) Prion-cured
K11 (YMD1180) Prion-cured
No cured strain available
Y12 (YMD1182) Prion-cured

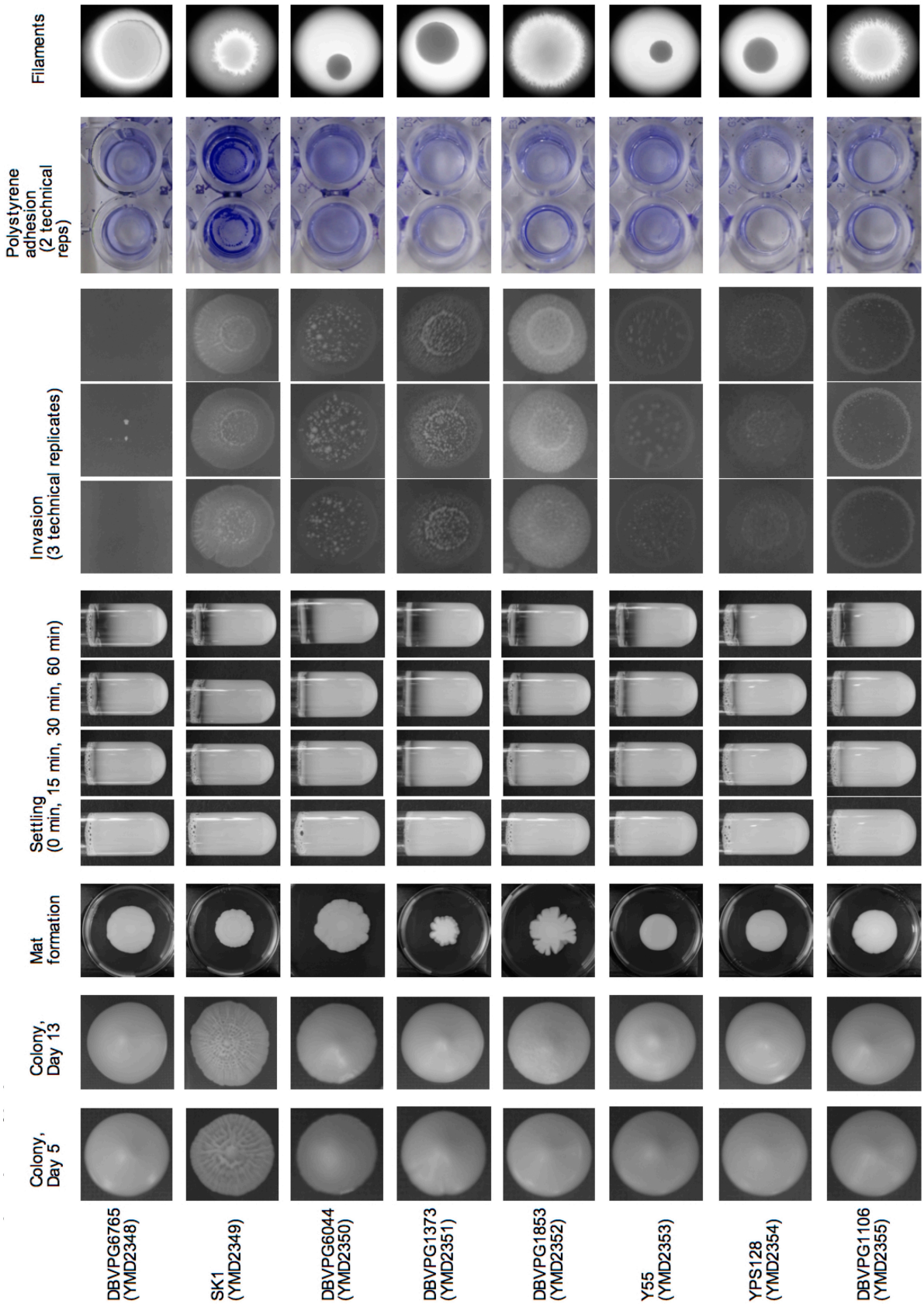

Colony, Day 13
Settling (0 min, 15 min, 30 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical replicates)
YS2 (YMD1184) Prion-cured
YS4 (YMD1186) Prion-cured
YS9 (YMD1188) Prion-cured
UWOPS83-787.3 (YMD1190) Prion-cured
UWOPS87-242.1 (YMD1192) Prion-cured
UWOPS05-217.3 (YMD1194) Prion-cured
UWOPS05-227.2 (YMD1196) Prion-cured
W303 (YMD1198) Prion-cured

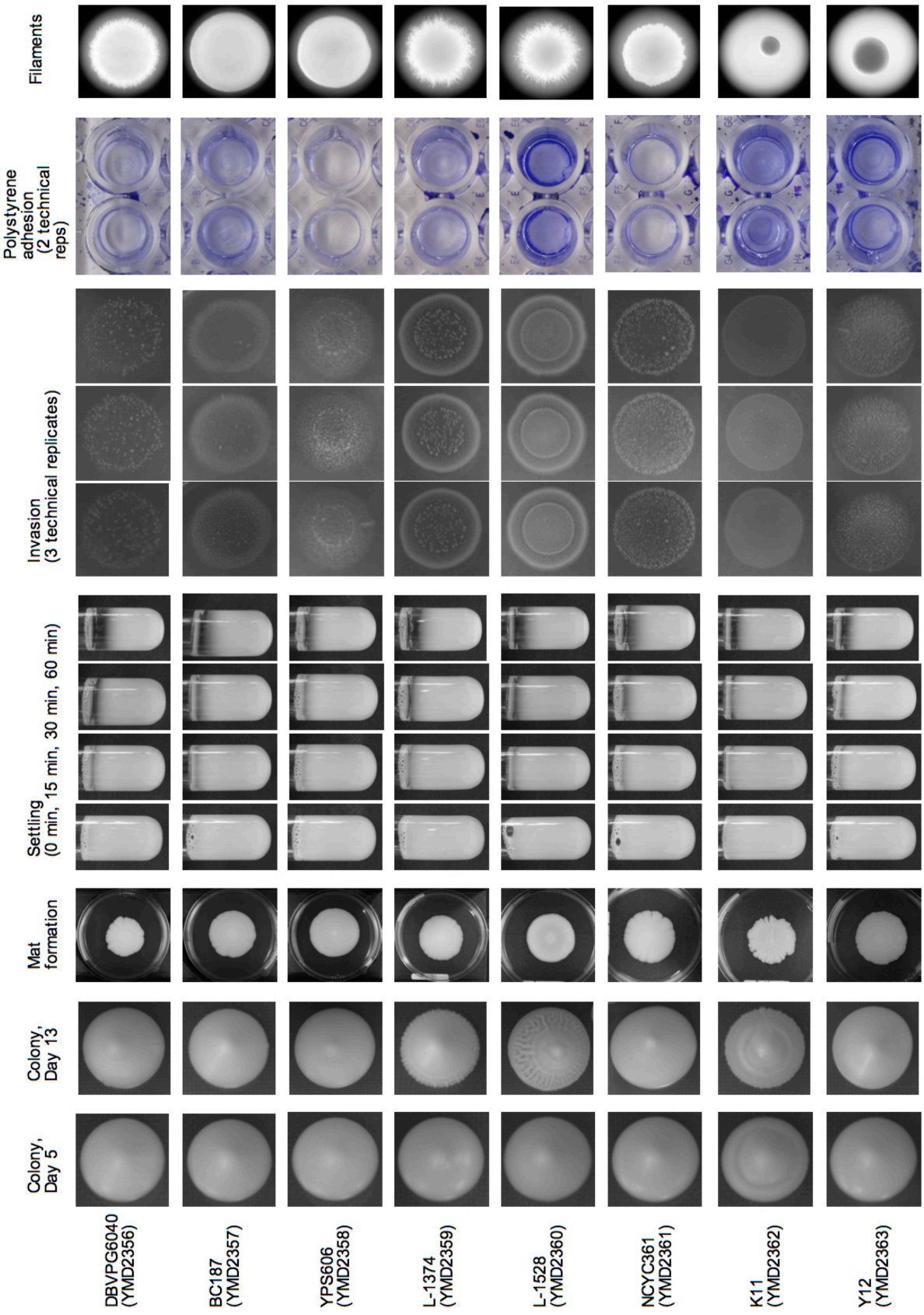

Colony, Day 13
Settling (0 min, 15 min, 30 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical replicates)
322134S (YMD1200) Prion-cured
378604X (YMD1202) Prion-cured
273614N (YMD1204) Prion-cured
YJM978 (YMD1206) Prion-cured
YJM981 (YMD1208) Prion-cured
YJM975 (YMD1210) Prion-cured
FY4 (DBY11069) Prion-cured

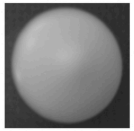
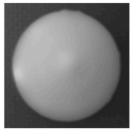
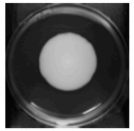
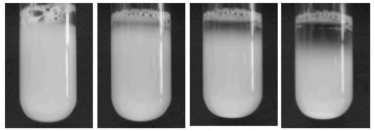
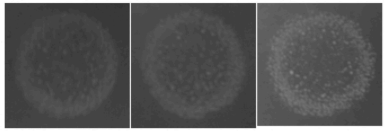
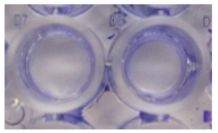
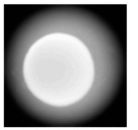
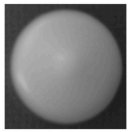
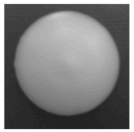
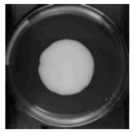
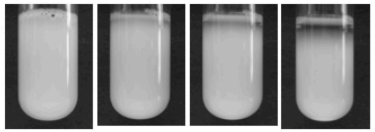
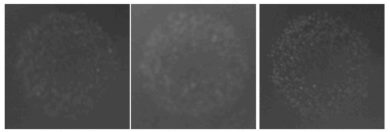
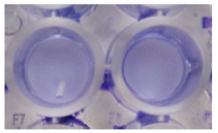
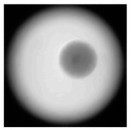
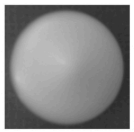
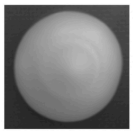
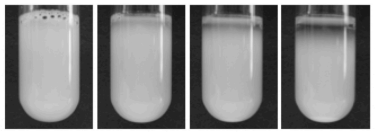
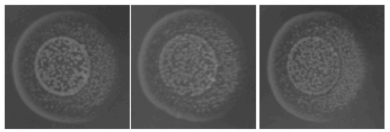
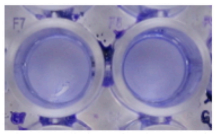
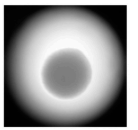
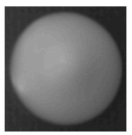
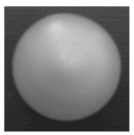
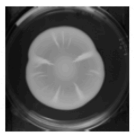
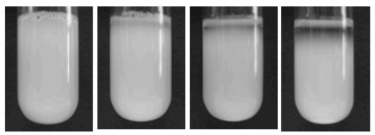
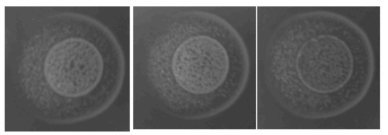
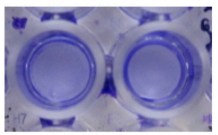
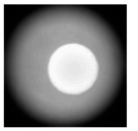


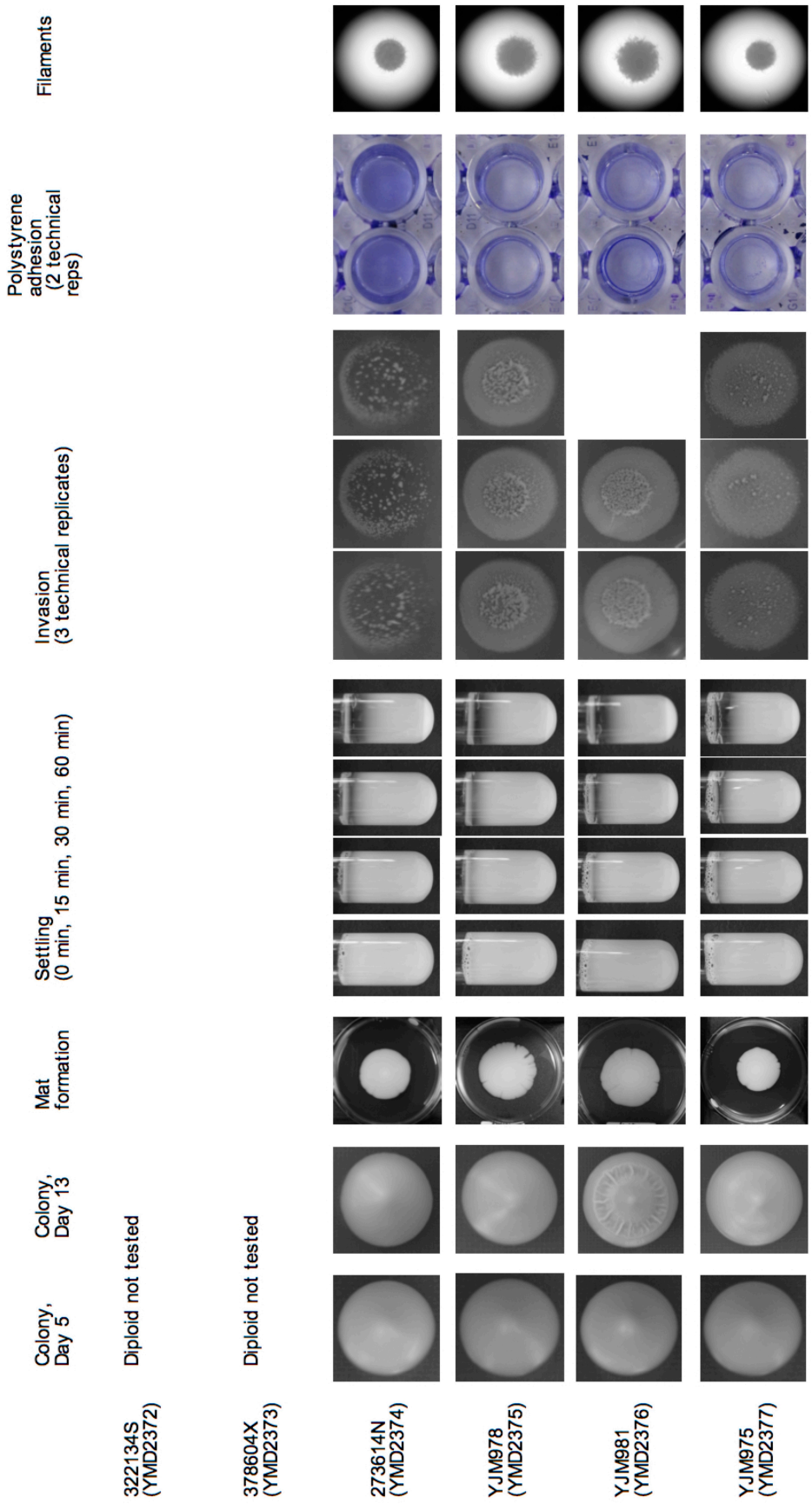
Full phenotypic panel for all 25 prion-cured haploid strains used in study. Strains are listed with their formal name and origin and are shown across five different phenotypes. Three technical replicates are shown for the invasion assay, photographed after 24h growth following washing on day 5. Two technical replicates are shown for the polystyrene adhesion assay. Pictured biofilms are fixed and stained with a 1% w/v crystal violet solution. Cured replicates are shown as the final strains in the panel.

Figure A.3. Complete diploid phenotypic panel



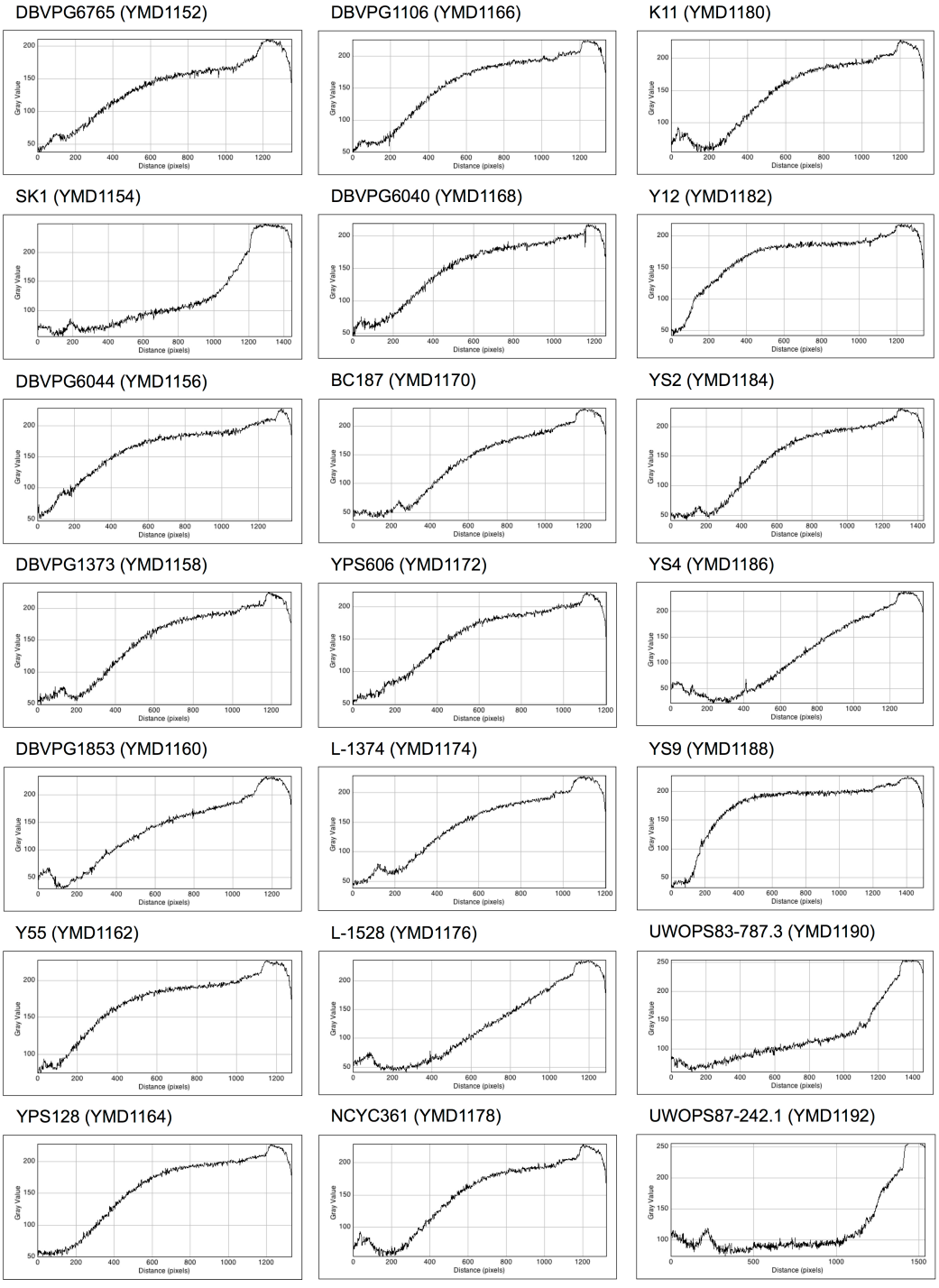

Colony, Day 5
Colony, Day 13
Mat formation
Settling (0 min, 15 min, 30 min, 60 min)
Invasion (3 technical replicates)
Polystyrene adhesion (2 technical reps)
Filaments
DBVPG6040 (YMD2356)
BC187 (YMD2357)
YPS606 (YMD2358)
L-1374 (YMD2359)
L-1528 (YMD2360)
NCYC361 (YMD2361)
K11 (YMD2362)
Y12 (YMD2363)

	Colony, Day 5	Colony, Day 13	Mat formation	Settling (0 min, 15 min, 30 min, 60 min)	Invasion (3 technical replicates)	Polystyrene adhesion (2 technical reps)	Filaments
YS2 (YMD2364)	Diploid not tested						
YS4 (YMD2365)	Diploid not tested						
YS9 (YMD2366)	Diploid not tested						
UWOPS83-787.3 (YMD2367)							
UWOPS87-242.1 (YMD2368)							
UWOPS05-217.3 (YMD2369)							
UWOPS05-227.2 (YMD2370)							
W303 (YMD2371)	Diploid not tested						

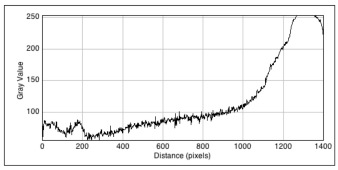


Full phenotypic panel for all 24 diploid strains used in study. Strains are listed with their formal name and origin and are shown across six different phenotypes. Three technical replicates are shown for the invasion assay, photographed after 24h growth following washing on day 5. Two technical replicates are shown for the polystyrene adhesion assay. Pictured biofilms are fixed and stained with a 1% w/v crystal violet solution.

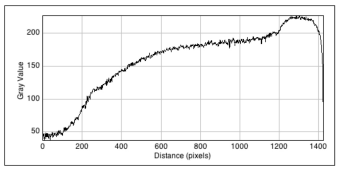
Figure A.4. Haploid settling plot profiles A representative plot profile generated by ImageJ is shown for each strain in one biological replicate of the haploid quantitative flocculation assay. Plot profiles show mean gray value along a line drawn on the 60-minute time point settling image from the meniscus of the culture to the bottom of the tube with corresponding pixel distance values on the x-axis of the plot.



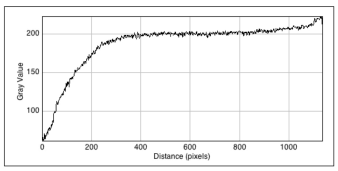
UWOPS05-227.2 (YMD1196)



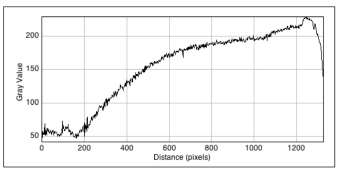
273614N (YMD1204)



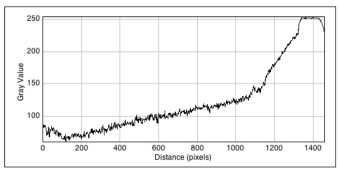
FY4 (DBY11069)



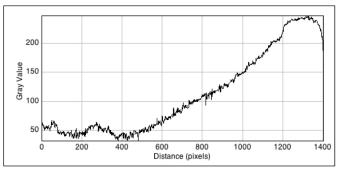
W303 (YMD1198)



YJM978 (YMD1206)



322134S (YMD1200)



YJM981 (YMD1208)

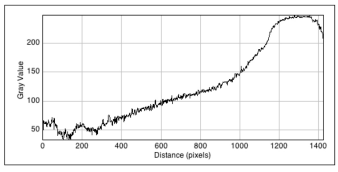
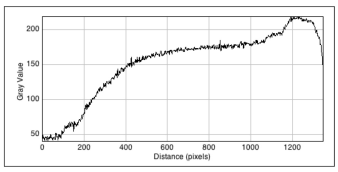
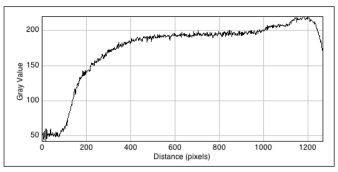


Figure A.5. Diploid settling plot profiles A representative plot profile generated by ImageJ is shown for each strain in the diploid quantitative flocculation assay. Plot profiles show mean gray value along a line drawn on the 60-minute time point settling image from the meniscus of the culture to the bottom of the tube with corresponding pixel distance values on the x-axis of the plot.

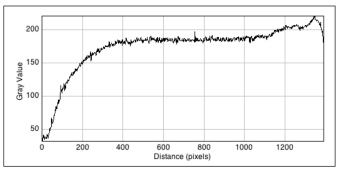
DBVPG6765 (YMD2348)



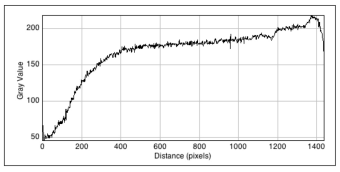
DBVPG1106 (YMD2355)



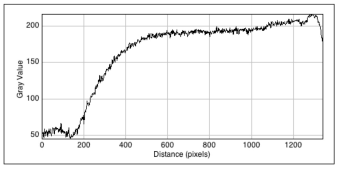
K11 (YMD2362)



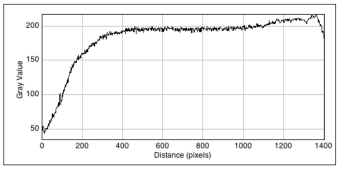
SK1 (YMD2349)



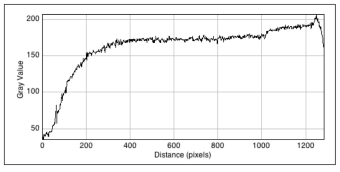
DBVPG6040 (YMD2356)



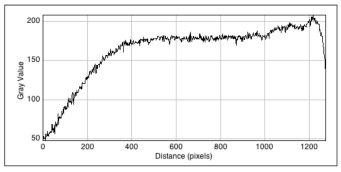
Y12 (YMD2363)



DBVPG6044 (YMD2350)



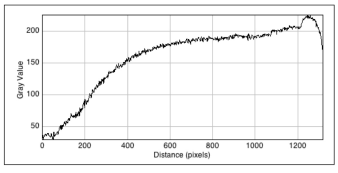
BC187 (YMD2357)



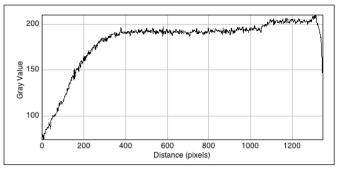
YS2 (YMD2364)

Diploid not tested

DBVPG1373 (YMD2351)



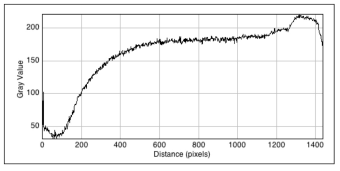
YPS606 (YMD2358)



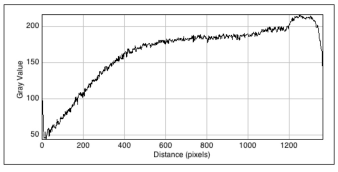
YS4 (YMD2365)

Diploid not tested

DBVPG1853 (YMD2352)



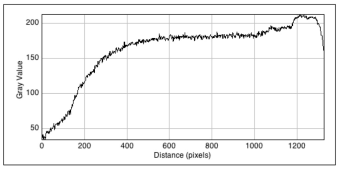
L-1374 (YMD2359)



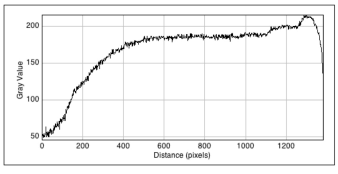
YS9 (YMD2366)

Diploid not tested

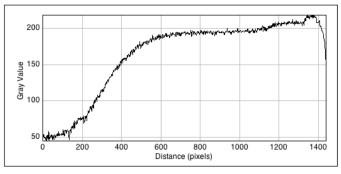
Y55 (YMD2353)



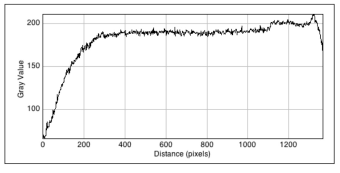
L-1528 (YMD2360)



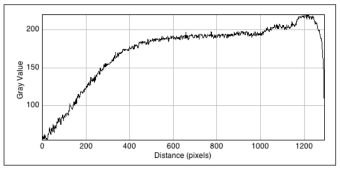
UWOPS83-787.3 (YMD2367)



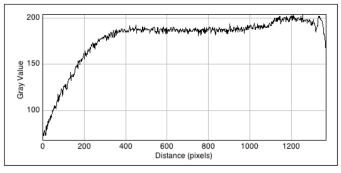
YPS128 (YMD2354)



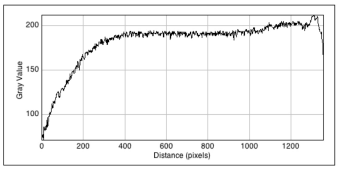
NCYC361 (YMD2361)



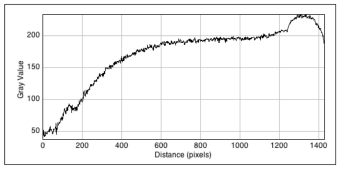
UWOPS87-242.1 (YMD2368)



UWOPS05-227.2 (YMD2370)



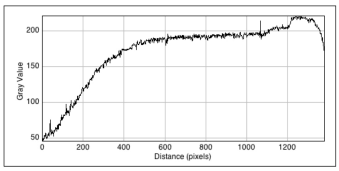
273614N (YMD2374)



W303 (YMD2371)

Diploid not tested

YJM978 (YMD2375)



322134S (YMD2372)

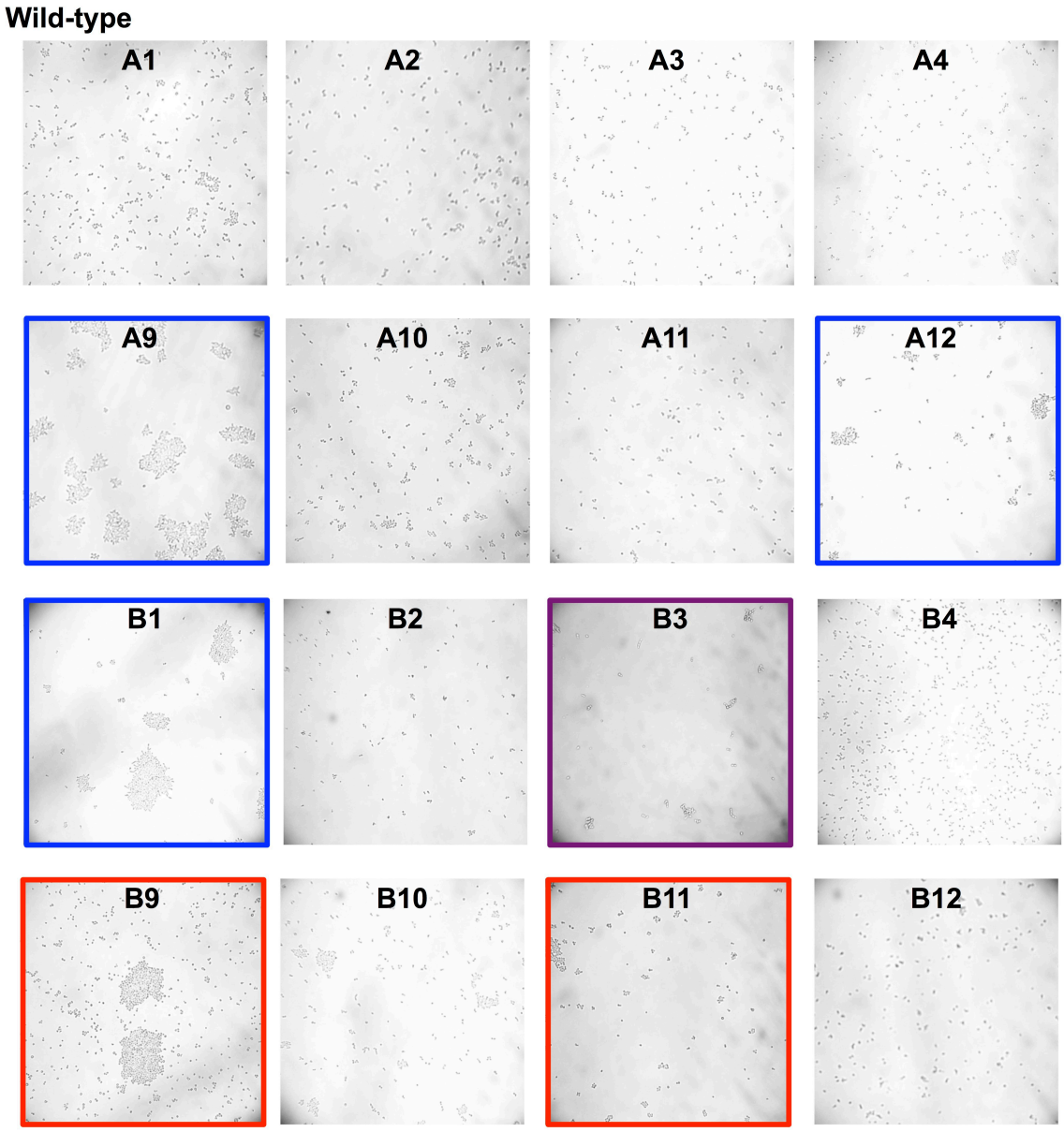
Diploid not tested

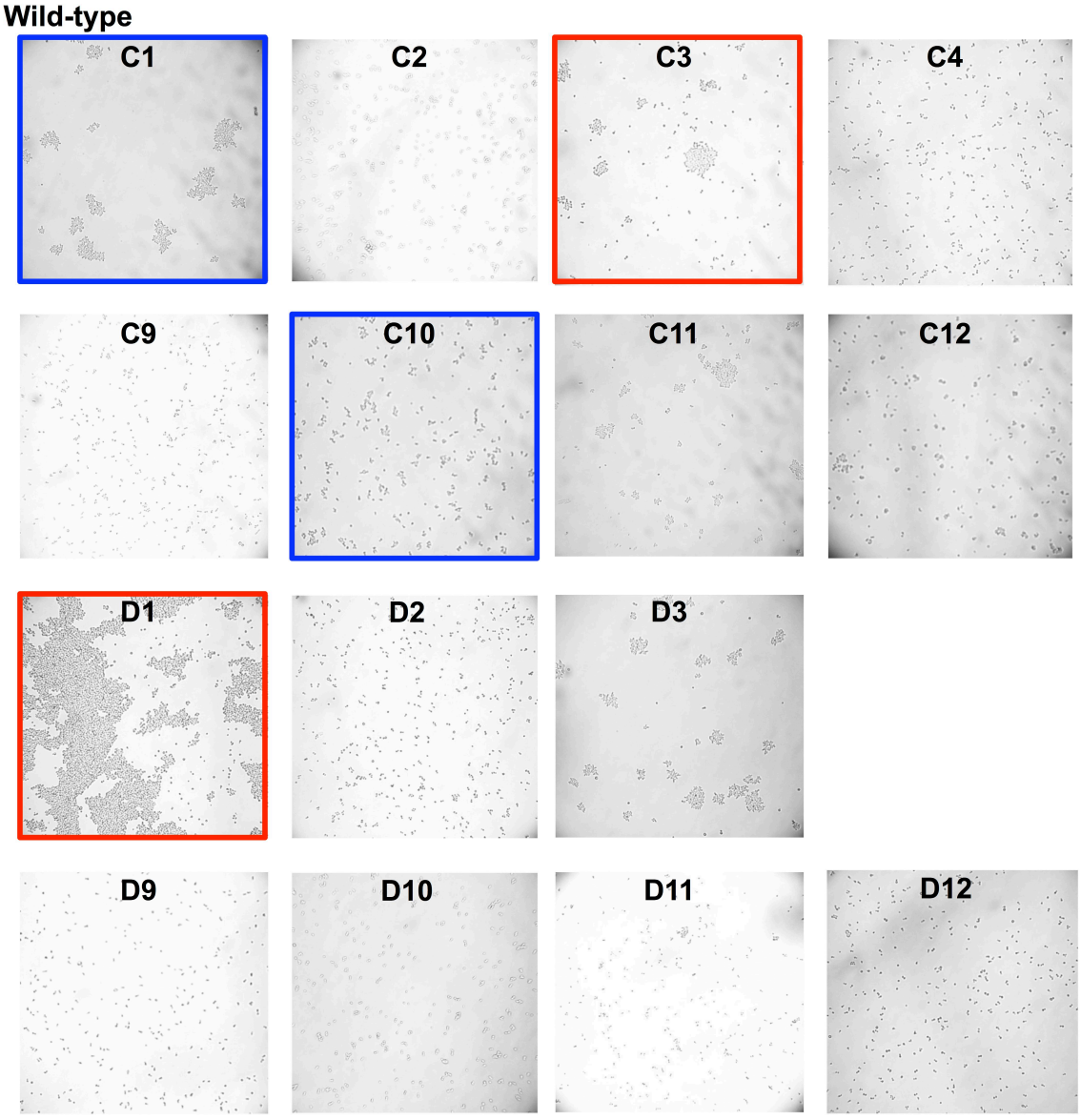
YJM981 (YMD2376)

Diploid not tested

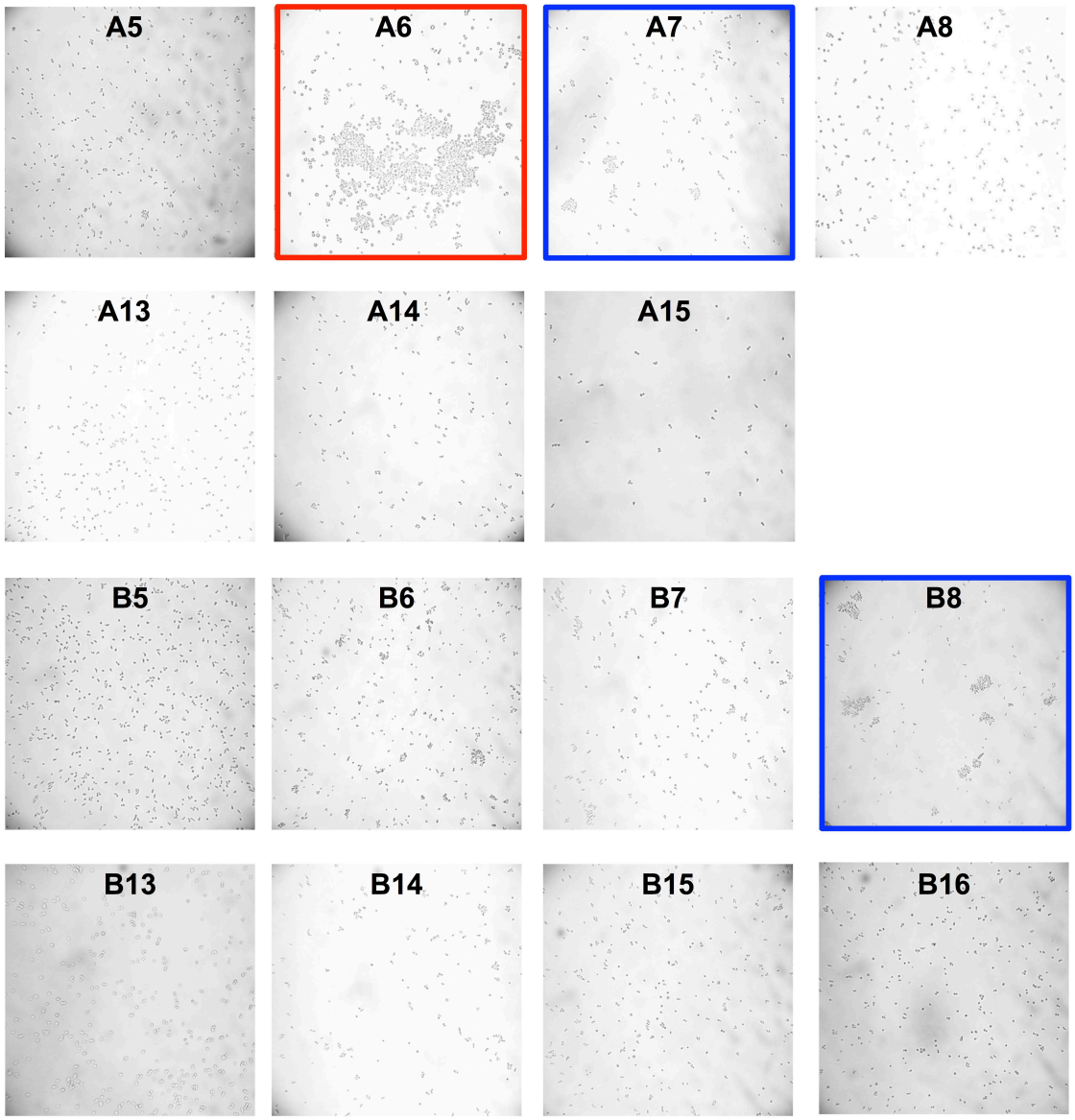
APPENDIX B

Figure B.1. Micrographs of evolved populations of wild-type and *flo1* knockout *S. cerevisiae* strains from the final experiment time point of the knockout evolution. Images are arranged to match the experimental setup in four 30°C heat blocks, A-D, receiving identical treatment. Red boxes surround images for vessels with a visible aggregation trait at the final time point, blue boxes surround images for vessels with mother-daughter separation defects recorded via microscopy, and a purple box surrounds the image for a vessel with a visible aggregation phenotype and possible separation defect. Vessels 1-4 and 9-12 in each block were inoculated with the wild-type strain, and vessels 5-8 and 13-16 contain the knockout strain. Missing images are for populations that were lost to contamination prior to the final time point. Micrographs received additional processing (grey scale conversion, 20% increase in brightness, 20% increase in contrast) to better highlight the phenotypes.

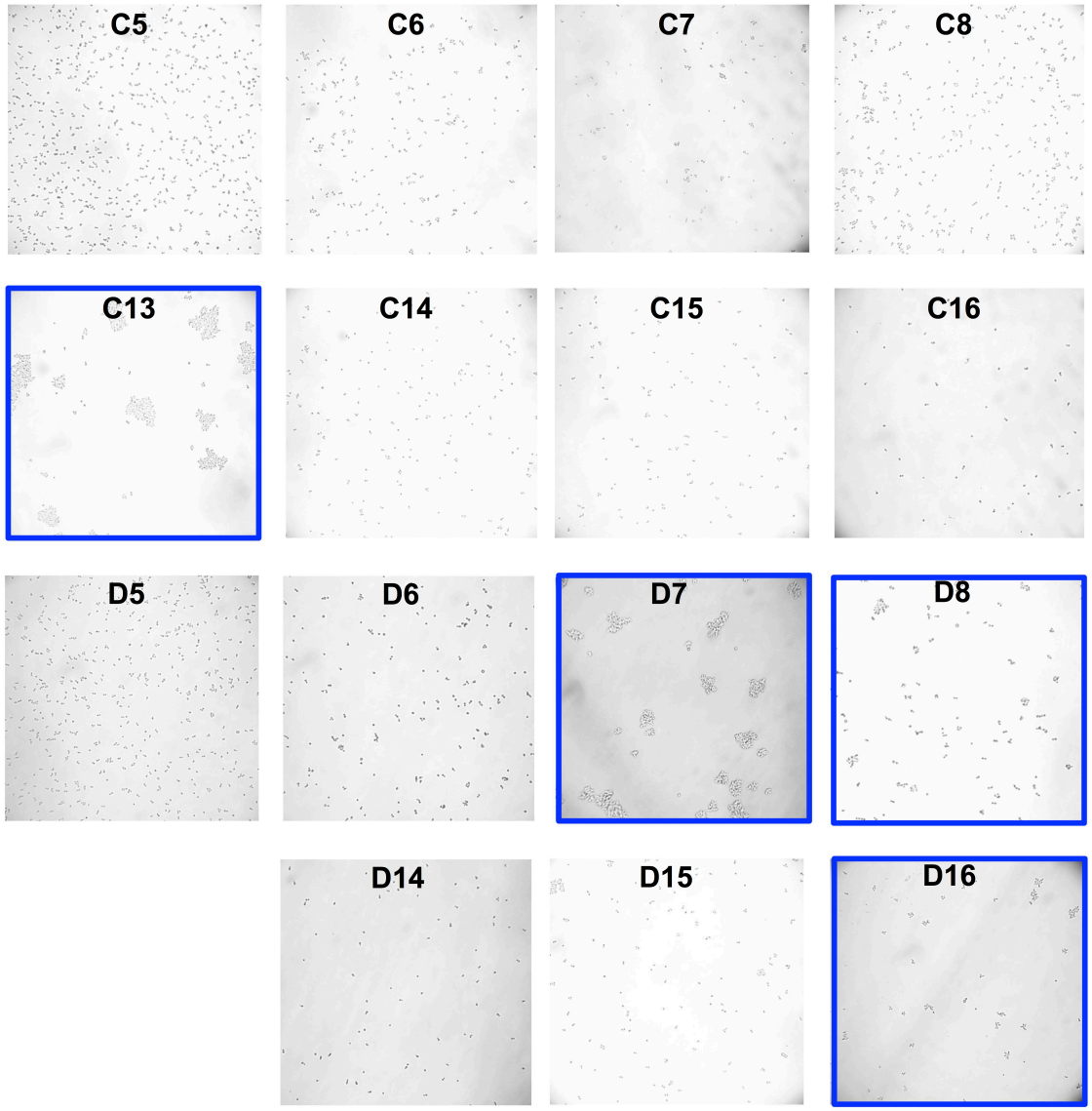



Wild-type
C1
C2
C3
C4
C9
C10
C11
C12
D1
D2
D3
D9
D10
D11
D12

***flo1* knockout**



***flo1* knockout**



VITA

Prior to matriculating at Stanford for her undergraduate studies, Elyse conducted research in solar modeling, for which she was recognized as an Intel Science Talent Search Finalist in 2006 and attended a graduate program in space weather modeling. She continued her research in atmospheric physics in the Geospace group at SRI International under Dr. Michael Nicolls and Dr. John Kelly, and co-authored a publication modeling atmospheric particle density phenomena. She received her B.S. in Biology from Stanford, where she conducted research in two groups: in Sports Medicine, studying patellar tendinopathy and gait biomechanics with Dr. Gordon Matheson; and in Genetics, studying vertebrate RNAi systems and immunoglobulin variable region arrangements with Dr. Andrew Fire and Dr. Scott Boyd. She joined the Genome Sciences doctoral program in the fall of 2011 and conducted her thesis work with Dr. Maitreya Dunham, examining the diversity of biofilm-related phenotypes in yeast natural isolates and the evolution of biofilm-related traits. She is continuing her study of yeast biofilm genetics as a senior fellow with Dr. Dunham.

During her time as a graduate student, Elyse engaged in numerous department and outreach activities, notably as the President of the Women in Genome Sciences advocacy organization from 2014-2016. Due to the importance of science fair research in her own education through middle and high school, she volunteers as a judge for the Northwest Association of Biomedical Research middle school essay competition and BioExpo fair, as well as for regional fairs in Oregon and Washington including the Washington State Science and Engineering Fair. Her recent doctorate made her eligible to judge at the international level and she has been selected as a Grand Awards judge for the International Science and Engineering Fair in 2017.

Publications:

E.A. Hope, C.J. Amorosi, A.W. Miller, K. Dang, C.S. Heil, M.J. Dunham. “Experimental evolution reveals favored adaptive routes to cell aggregation in yeast.” *Genetics*, under review.

E.A. Hope and M.J. Dunham. “Ploidy-regulated variation in biofilm-related phenotypes in natural isolates of Saccharomyces cerevisiae.” *G3: Genes| Genomes| Genetics* 4 (9), 1773-1786.

M.J. Nicolls, C.J. Heinselman, **E.A. Hope**, S. Ranjan, M.C. Kelley, J.D. Kelly. “Imaging of Polar Mesosphere Summer Echoes with the 450-MHz Poker Flat Advanced Modular Incoherent Scatter Radar", *Geophysical Research Letters*, VOL. 34, L20102, doi:10/29/2007