The Genetic Architecture of Biofilm Formation in a Clinical Isolate of *Saccharomyces cerevisiae*

Joshua A. Granek, Debra Murray, Ömür Kayıkçı, Paul M. Magwene

*Department of Biology and Center for Systems Biology, Duke University, Durham, NC, 27708*

Genbank accession numbers: JN974464 and JN974465
Running Title: Genetic Architecture of Yeast Biofilms

Keywords: Biofilm, Colony Morphology, Bulk Segregant Analysis, cAMP-PKA, Natural Variation

Corresponding author:
Paul M. Magwene
Department of Biology,
Duke University,
P.O. Box 90338
Durham, NC 27708
USA
Phone: 919-613-8159
Fax: 919-660-7293
Email: paul.magwene@duke.edu
ABSTRACT

Biofilms are microbial communities that form on surfaces. They are the primary form of microbial growth in nature and can have detrimental impacts on human health. Some strains of the budding yeast Saccharomyces cerevisiae form colony biofilms, and there is substantial variation in colony architecture between biofilm-forming strains. To identify the genetic basis of biofilm variation, we developed a novel version of quantitative trait locus mapping, which leverages cryptic variation in a clinical isolate of S. cerevisiae. We mapped thirteen loci linked to heterogeneity in biofilm architecture and identified the gene most closely associated with each locus. Of these candidate genes, six are members of the cyclic AMP – protein kinase A pathway, an evolutionarily conserved cell signaling network. Principal among these is CYR1 which encodes the enzyme that catalyzes production of cAMP. Through a combination of gene expression measurements, cell signaling assays, and gene over-expression, we determined the functional effects of allelic variation at CYR1. We found that increased pathway activity resulting from protein coding and expression variation of CYR1 enhances the formation of colony biofilms. Four other candidate genes encode kinases and transcription factors that are targets of this pathway. The protein products of several of these genes together regulate expression of the sixth candidate, FLO11, which encodes a cell adhesion protein. Our results indicate that epistatic interactions between alleles with both positive and negative effects on cyclic AMP – protein kinase A signaling underlie much of the architectural variation we observe in colony biofilms. It is also among the first to demonstrate genetic variation acting at multiple levels of an integrated signaling and regulatory network. Based on these results, we propose a mechanis-
tic model that relates genetic variation to gene network function and phenotypic outcomes.

ABBREVIATIONS

QTL, quantitative trait locus; BSA-seq, Bulk Segregant Analysis via high-throughput sequencing; cAMP-PKA, cyclic AMP-protein kinase A; YEPLD, yeast extract peptone low-dextrose
INTRODUCTION

Biofilms are complex, surface-adherent communities of microbes. They are ubiquitous in nature and in the human environment (López et al., 2010; Donlan and Costerton, 2002). Not only can they be life-threatening when they form in the human body (Hall-Stoodley et al., 2004), but biofilms also create problems when they form on human-made structures, ranging from merely annoying (shower curtains) to hazardous (nuclear power plants) (Flemming, 2002). Because of the frequency with which we interact with biofilms, characterizing the cellular mechanisms that regulate them and understanding the genetic basis of their variation are of great interest.

The ability to form biofilms is a multifactorial trait, resulting from the interactions of several lower-level properties, including cell-cell and cell-substrate adhesion, production of extracellular matrix, and spatial heterogeneity in cell morphology, growth, and division. Recent studies have demonstrated that the budding yeast *Saccharomyces cerevisiae* forms microbial communities that have these biofilm characteristics (Kuthan et al., 2003; Váchová et al., 2011). These characteristics are present in mats that form on soft agar (Reynolds and Fink, 2001), and in complex colony biofilms (sometimes referred to as “fluffy colonies”) that are distinguished from non-biofilm colonies (also described as simple or smooth colonies) by intricate, organized, and strain specific architecture (Figure S1) (Granek and Magwene, 2010; Palková and Váchová, 2006; Honigberg, 2011). These complex colony biofilms are also formed by the more prevalent fungal pathogens *Candida albicans* (Hall et al., 2010) and *Cryptococcus neoformans* (Goldman et al., 1998), where they are associated with pathogenicity.
There is substantial variation among natural isolates of *S. cerevisiae* in the propensity to form biofilms, the architectures produced, and the effect of environmental signals on biofilm phenotypes (i.e. genotype-by-environment interactions) (Palková, 2004; Granek and Magwene, 2010). The nature of this variation suggests that biofilm formation is a genetically complex trait, and that allelic variation at multiple loci contributes to phenotypic heterogeneity. Morphological complexity seems to be an integral character of fungal biofilms; it is a high-level property arising from cell-cell interaction, cell-substrate adhesion, and extracellular matrix production, the same properties essential to biofilm formation, so it is an excellent characteristic to assay for biofilm behavior. The relative complexity of a yeast colony therefore serves as a readily assayed indicator of biofilm formation.

In this study we exploit cryptic variation within a clinical isolate of *S. cerevisiae*, resulting from genomic heterozygosity, to map loci that contribute to variation in colony biofilm attributes. YJM311 is a homothallic diploid strain that was collected from the bile tube of a patient in the San Francisco Bay area in 1991 (McCusker *et al.*, 1994b; Clemons *et al.*, 1994). This strain displays several pathogenic traits common to clinical isolates, including virulence in mice, pseudohyphal growth, and resistance to high temperature (Clemons *et al.*, 1994; McCusker *et al.*, 1994b). YJM311 produces complex colony biofilms when grown under conditions of glucose limitation (Figure 1A; (Granek and Magwene, 2010)). The homozygous diploid segregants generated directly from YJM311 exhibit great diversity in colony architecture ranging from simple non-biofilm morphologies to complex biofilm morphologies exhibiting a wide variety of morphotypes (Figure 1). These segregants also vary in their adherence to plastic (Figure S2), a feature of biofilms that form on indwelling medical devices.

Here we describe the application of Bulk Segregant Analysis via high-throughput
sequencing (BSA-seq; Schneeberger et al., 2009; Ehrenreich et al., 2010; Michelmore et al., 1991; Brauer et al., 2006; Magwene et al., 2011b) to identify loci that contribute to variation in yeast colony biofilm formation. We developed a novel version of this method, which is distinct from standard BSA-seq in two ways. First, we used the natural heterozygosity within YJM311 as the basis for our mapping population. The genome of YJM311 is highly heterozygous (more than 40,000 heterozygous SNPs, representing more than 0.35% of the genome), and genetically distinct from the genome of the standard reference strain S288c (Magwene et al., 2011a). We produced an enormously diverse population of homozygous diploid segregants from this heterozygous diploid through random sporulation. Since this strain is homothallic, the haploids generated by sporulation spontaneously self-mated to form homozygous diploids with a unique combination of alleles at the loci that were heterozygous in the parent. Second, we employed multiple bulk comparisons. BSA typically involves comparison of allele frequencies between a single pair of bulks; for example between a complex and simple bulk representing the extremes of the phenotypic distribution. In this study we used three segregant bulks – representing the two tails of the phenotypic distribution plus a set of intermediate segregants. As we demonstrate below, this three bulk approach allowed us to tease apart multiple overlapping quantitative trait loci (QTLs) that would otherwise have been impossible to disambiguate had we used only the extreme bulks.

Using BSA-seq we mapped thirteen colony biofilm QTLs and identified a candidate gene associated with each QTL peak. Of these thirteen genes, six contribute to or are targets of the cyclic AMP – Protein kinase A (cAMP-PKA) pathway, a signal transduction cascade conserved throughout the animal and fungal kingdoms (Taylor et al., 1990; Francis and Corbin, 1994). One of these candidates, CYR1, en-
codes adenylate cyclase, the enzyme that catalyzes the conversion of ATP to the second messenger cAMP. We characterized the mechanisms of variation at CYR1, and our results suggest that there is functional variation both at the coding sequence level and at the level of gene regulation. Among the other candidate genes identified in our mapping study are transcription factors (SFL1, FLO8, and MSN2) and a kinase (YAK1), all of which participate in a regulatory network downstream of the cAMP-PKA pathway. This network controls several nutrient response mechanisms, including expression of another candidate gene, FLO11, which encodes a flocculin, a protein that mediates adhesion of yeast cells to each other, and to surfaces (Guo et al., 2000; Verstrepen et al., 2004). FLO11 is directly regulated by SFL1 and FLO8, (Pan and Heitman, 2002) and indirectly by YAK1 (Zhang et al., 2001).

Variation in both cAMP levels and the expression of several of the candidate genes is strongly correlated with variation in colony biofilm architecture. To our knowledge, this study is the first to dissect the genetic architecture of fungal biofilms and to propose a mechanistic model of how genetic variation acting at multiple nodes across a gene network influences biofilm phenotypes. As one of the few observations of segregating variation acting at multiple levels of a genetic network (Gerke et al., 2009), these results provides a novel insight into the role of network variation on complex traits. Because the cAMP-PKA pathway is an pleiotropic regulator of many developmental switches in fungi (Granek et al., 2011; McDonough and Rodriguez, 2012), including many that are important for pathogenesis (Hall et al., 2010; Yi et al., 2011), our findings will serve as an important guide for future efforts aimed at understanding natural variation in fungal development and morphogenesis.
MATERIALS AND METHODS

Media, Strains, Plasmids, and Primers

YPD (Sherman, 2002), SC (Sherman, 2002), Sporulation Media (Sherman, 2002), and YEPLD (Granek and Magwene, 2010) were made as described. G418 selective media was made with 200 mg/L Geneticin. Tetracycline-regulating “Dox” media contained 10 mg/L doxycycline, except where otherwise indicated. YPD+cAMP and YPD+water media was prepared by spreading 667 µL of 30 mM cAMP (A9501-1G, Sigma Aldrich) yielding a final concentration of 1 mM cAMP, or 667 µL sterile water on the surface of 20 mL YPD agar petri dishes, and allowing it to dry overnight.

Strains used are detailed in Table 1. BY4743 was used as a host for plasmid construction, and all YJM311 segregants were derived by random sporulation, they are diploid due to self-mating. The “assay subset” of segregants that was used in experiments beyond the BSA-seq, consisted of thirty-five segregants randomly selected from the simple pool, and thirty-five segregants randomly selected from the complex pool. Plasmids are listed in Table 2, and PCR primer information is in File S1.

Bulk Segregant Analysis by Pooled Genome Sequencing

BSA-seq was used to identify the loci contributing to the morphological variation. We divided segregants of YJM311 into one of three groups: simple, intermediate, and complex. From each group, a subset of 288 segregants (three 96-well plates) was selected to generate a bulk for pooled sequencing.
Segregants were generated by random sporulation following the procedure of Goddard et al. 2005 (Goddard et al., 2005) with the following changes. Cells grown to saturation in YPD at 30 °C overnight were washed, resuspended in sterile water, plated on Sporulation Media and incubated at room temperature for nine days. Sporulated cells were suspended in sterile water by gently scraping the surface of the agar with a sterile cell scraper. The cell suspension was transferred to a 14 mL plastic culture tube, washed two times with sterile water and resuspend in 4 mL of Outcrossing Solution (Goddard et al., 2005). After sonication, cells were washed two times in sterile water, and resuspended in 1.5 mL sterile 0.01% SDS.

The random spore suspension was plated on YEPLD agar and incubated at 30 °C. Once grown, the morphology of colonies was judged by visual inspection for assignment to a pool, they were picked, suspended in 100 µL YPD in a 96-well plate well, and incubated overnight at 30 °C. Segregants were replicated from the 96-well plate to a YEPLD agar omnitray (Nunc #242811) using a 96 Solid Pin Multi-Blot Replicator (V&P Scientific #VP408FP6), and grown to confirm morphology.

To determine allele frequencies of all SNPs in the pools, each segregant was grown individually in a 96-well plate in YPD at 30 °C for two days. Individual cultures from each 96-well plate were aspirated and pooled in a pre-weighed 50 mL conical tube. Wells were washed with water to gather adherent cells and these suspensions were added to the pool. The cell suspension was centrifuged to pellet, the supernatant was removed, the pellet was resuspended in 4 mL TE buffer, pelleted again, the supernatant was again removed, the conical tube was weighed to determine the total cell mass, and frozen at −80 °C. DNA was extracted from the pools using anion-exchange gravity flow columns (Genomic-Tip, Qiagen). To construct the genomic libraries, we used approximately 5 µg of material from each pool and followed Illumina Sample Preparation Kit instructions. Libraries were quantified.
using Experion electrophoresis (Bio-Rad) and Qubit fluorometer (Invitrogen), and then processed on the Illumina GAII Genome Analyzer at Duke University’s DNA Sequencing Facility.

The simple and complex pools were sequenced using two Illumina flow-cell lanes each, while the intermediate pool was sequenced using a single lane. The approximate total number of reads for each pool were: 30.3M for simple, 14.5M for intermediate, 31.7M for complex. Short-reads were mapped to the standard S. cerevisiae reference genome as described above. The average depth of coverage for each pool, based on mapped reads were: 120× for simple, 55× for intermediate; 126× for complex.

We sequenced the heterozygous diploid strain YJM311 using 50bp reads on an Illumina GAII Genome Analyzer at the Duke University DNA Sequencing Facility. Sequencing yielded approximately 15M short reads, which were mapped to the S288c standard S. cerevisiae reference genome (obtained from the Saccharomyces Genome Database, January 2010), using the short-read mapping software BWA, version 0.5.0 (Li and Durbin, 2009). SNP calls were made using the “pileup” and “varFilters” options of SAMtools, version 0.1.7 (Li et al., 2009). In total 44,076 high quality heterozygous sites were identified in strain YJM311. This number differs from the number of heterozygous sites reported in Magwene et al. (2011) (Magwene et al., 2011a), which was based on 36bp reads and used a more conservative criterion for identifying heterozygous sites based on the intersection of SNP calls from two different short-read mapping algorithms.

Sequence data was processed to identify loci linked to colony morphology using a combination of publicly available packages and custom Python scripts. At each site identified as heterozygous in YJM311 we estimated the counts of each nucleotide allele in both YJM311 and the segregant pools based on “pileup” files.
generated via SAMtools and custom scripts. To map loci that contribute to variation in colony morphology we carried out three pairs of comparisons between pools (simple vs. intermediate, intermediate vs. complex, simple vs. complex).

Similarly, to identify deleterious loci we carried out three comparisons between each pool and YJM311 (simple vs. YJM311, intermediate vs. YJM311, and complex vs. YJM311). There are two sources of variation in the BSA-seq analyses that affect allele counts in pools. One is variation due to the sampling of segregants when forming the pools. The second is variation that arises due to the measurement technique itself (library preparation, sequencing coverage, post-sequencing alignment of reads, etc). Magwene et al. (2011b) describe a statistical and analytical framework that accounts for these two sources of variation based on smoothed G-statistics. For each pairwise comparison we used this smoothed G-statistic, $G'$, to characterize the data at each site based on a smoothing window size $W = 33,750$ Kb ($\sim 25$ cM; see Magwene et al. (2011b) for detailed discussion of $G'$ and the choice of smoothing window).

We used a false discovery rate threshold of 0.01 to identify QTL regions, and developed a peak calling algorithm to identify peaks within these regions. Briefly, a QTL region was defined as a continuous run of at least ten SNPs, spanning at least 10Kb, where the $G'$ statistic exceeded the false discovery rate threshold. Peaks were identified as contiguous subregions such that $G'_i \geq 0.90G'_{\text{max}}$, where $G'_{\text{max}}$ is the site that has the largest $G'$ within the QTL region. Details of statistically significant peaks are provided in Supplementary File S2. For each QTL peak we identified candidate genes based on criteria such as the number of variable sites, the number of predicted non-synonymous substitutions, and functional annotation.
Colony Morphology Quantitation

Colony morphology was quantitated using criteria described previously (Granek and Magwene, 2010). Colonies were grown on YEPLD omnitrays at 30 °C and imaged by flatbed scanner (Epson Expression 10000XL). Images were then scored by visual inspection; scores were assigned on a scale from one to five to one decimal place, with one indicating completely simple, non-biofilm phenotypes and five indicating highly complex biofilm colony morphologies. Colony morphology scores in Figure 1 are the median of at least four replicate colonies scored on day 4 (for 16 of 288 segregants only two replicate colonies were used). Scores in Figure S4 are the median of six replicate colonies scored on day 7.

Intracellular cAMP

Intracellular cAMP concentrations in the assay subset were determined by growing segregants on YEPLD for three days at 30 °C. Individual colonies were then scraped from plates and suspended in 500 µL 1 M n-butanol saturated formic acid to fix, lysed by four freeze-thaw cycles in dry ice. The formic acid was then evaporated using a Speed Vac Concentrator (Savant Instruments, Hickville, NY). cAMP levels were quantified following the Non-Acetylation EIA protocol for the cAMP Biotrak EIA kit (RPN225, GE Healthcare).

Gene Expression

To quantitate gene expression in the assay subset, segregants were grown on YEPLD for three days at 30 °C. Individual colonies were then scraped from plates and suspended in ice cold water. Cells were pelleted by centrifugation, super-
natant was removed, pellets were flash frozen in liquid nitrogen, and stored at −80 °C. RNA was extracted from cell pellets by the hot acid phenol method, as previously described (Collart and Oliviero, 2001), except that Phase Lock Gel Light 1.5 ml tubes (5 PRIME, Gaithersburg, MD) were used in place of standard microcentrifuge tubes during the phenol and chloroform extract steps. Quantitation of gene expression was conducted by NanoString Technologies (Seattle, WA) using their nCounter platform. All gene expression analysis utilized data normalized by standard spike-in controls (Geiss et al., 2008).

**Major Locus Genotyping**

Based on the BSA-seq results, we identified SNPs in three regions of the genome with the greatest potential for significant allelic differences: CYR1 and FLO11. We confirmed the SNPs across the regions of interest by Sanger sequencing and selected SNPs at either end to assay by HRM (High Resolution Melt) analysis. HRM primers were designed such that they had melting temperatures of 56 °C, that pairs had similar melting temperatures, and would amplify 100–150bp fragments surrounding each SNP. To ensure a high degree of specificity, we BLASTed each primer against the yeast genome and verified it by PCR prior to use. In addition, we checked the folding characteristics of the primers using DINAMelt (Markham and Zuker, 2005). The primers used are described in Supplementary File S1. The and SNPs assayed are detailed in Table S1; coordinates in this table are based on Saccharomyces Genome Database release r64, (release date: February 4, 2011) (Cherry et al., 1997). Sanger sequencing was performed by the Duke IGSP Genome Sequencing & Analysis Core Resource on a 3730xl DNA Analyzer (Applied Biosystems).
Genomic DNA templates for HRM were isolated from segregants using the DNeasy Blood & Tissue Kit, following the manufacturer’s supplementary yeast protocol DY13 (Qiagen), except 20 µL of 10 mg/mL Zymolyase was used instead of lyticase. All extracts were resuspended in 200 µL of the provided AE buffer and then quantified with an ND-1000 Spectrophotometer (NanoDrop) in order to normalize concentrations. Dilution steps were carried out using the same AE buffer.

All HRM assays were run on the Rotor-Gene 6000 (Qiagen) using EvaGreen fluorescent dye (Type-it HRM kit, Qiagen) and 30 ng of DNA. In each run we included six samples with known genotypes as controls, two of each allele and two heterozygous for the SNP in question. All runs were conducted in triplicate. For more accurate loading of the 10µl reactions into the Rotor-Disc 100 microtubes, we used the CAS-1200 Automated Sample Setup Robotic (Corbett). Cycling protocols followed the Type-it HRM kit instructions for SNP analysis, using 0.1 °C increments in the initial range of 65 – 95 °C, with a 2 second hold at each step.

HRM data were analyzed using the Rotor-Gene 6000 software. We visually inspected the real-time amplification plots to verify a successful run. Because poorly performing samples can negatively impact the overall HRM analysis, we removed any outliers prior to data analysis. This includes samples that amplified late (Ct values over 30), failed to amplify, or showed markedly lower end-point fluorescence levels than other samples. We called the genotypes of the samples using our known controls to designate three genotypes (both alleles and the heterozygous state). Our standard for acceptance of assigned genotypes was 90% confidence for at least two of three replicates. Low confidence calls, heterozygous calls, or conflicts between run (usually occurring together) were confirmed by Sanger sequencing. At each locus, the allele carried by the majority of complex segregants was designated “complex” (C), and the alternate allele was designated “simple”
YJM311 is homothallic, so the default expectation is that each segregant is diploid, arising from a single spore that underwent a mating-type switch after the first cell division, followed by self-mating. Genotyping indicated that three segregants (s32, s36, and s49) were heterozygous at one or more of the loci assayed. These heterozygotes likely resulted from inter- or intra-tetrad mating. It is possible that the heterozygotes arose from diploids that survived the random sporulation procedure. This is unlikely, since no colonies grew appeared from the random sporulation negative control, a culture of YJM311 grown to saturation in liquid YPD, and processed in parallel with the sporulated culture. These three segregants were excluded from further analyses.

**Calculation of Variance Explained**

Calculation of the variance in cAMP concentration, *CYR1* expression, and *FLO11* expression explained by the major loci was done for each using the `aov` function of R (R Development Core Team, 2011) to fit the model described by equation 1. The variance explained by each locus was calculated as the locus sum of squares divided by the total sum of squares.

\[
\text{Trait} = \text{CYR1.geno} + \text{FLO11.geno} + \text{Error} \quad (1)
\]

**CYR1 Overexpression**

The tetracycline-regulatable *lacZ* (pJG117) expression plasmid was derived from pCM179, and the *CYR1-S* (pJG118) and *CYR1-C* (pJG121) expression plasmids were derived from pCM190 (Garí *et al.*, 1997). The plasmids were constructed us-
ing a strategy based on the DNA assembler technique (Shao et al., 2009), which takes advantage of homologous recombination in yeast to assemble multiple DNA fragments with shared homologous regions into intact plasmids. In all three plasmids, the original URA3 marker was replaced by KanMX4, a drug-selectable marker, so they could be used in the strains studied here, which carry no auxotrophies. The KanMX4 marker was amplified from pFA6-KanMX4 (Wach et al., 1994), with FideliTaq (USB) using primers pcm_kan_up and pcm190_kan_dn, which have 50bp 5’-flanks homologous to the URA3 flanks in pCM179 and pCM190. To construct pJG118 and pJG121, genomic DNA was isolated from segregants s4 and s33 using the DNeasy Blood & Tissue Kit, following the manufacturer’s supplementary yeast protocol DY13 (Qiagen). This genomic DNA was used as template for PCR to amplify the CYR1-S and CYR1-C alleles, respectively. The primers used, pcm_cyr1_up and pcm_cyr1_dn, have 50bp 5’-flanks homologous to the multiple-cloning site in pCM190. The cycling program and reaction conditions were optimized to produce sufficient product for cloning while minimizing the risk of mutations. 25 µL PCR reactions consisted of Phusion Buffer GC, 200 µM each dNTP, 0.5 µM each primers, 100-150 ng genomic DNA, 5% DMSO, 2.1 mM MgCl2 (total), 0.75U Phusion Polymerase. The cycling program was 2 min at 98°C; 25 cycles of: 15 s at 98°C, 30 s at 59°C, 190 s at 72°C; with a final extension of 10 min at 72°C, and was carried out in a S1000 Thermal Cycler (Bio-Rad). The PCR products were gel-purified and extracted using the QIAquick Gel Extraction Kit (Qiagen).

PCR products and plasmid were co-transformed into strain BY4743 using the lithium acetate method (Gietz and Schiestl, 2007). pJG117 was generated by co-transforming pCM179 with the KanMX4 PCR product. pJG118 and pJG121 were assembled by co-transforming the CYR1-S and CYR1-C PCR products, respectively, the KanMX4 PCR product, and pCM190 digested with EcoRV and BamHI.
to cut within URA3 and the multiple-cloning site. Colonies carrying the plasmids were selected by growth on YPD+G418+Dox. Doxycycline was incorporated to block expression of CYR1 and lacZ, in order to avoid selecting against functional plasmids, which might have a higher fitness cost than non-functional plasmids. Plasmids were rescued from yeast by QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer protocol for yeast, transformed into *Escherichia coli* (ElectroSHOX Competent Cells, Bioline), and purified by QIAprep Spin Miniprep Kit (Qiagen) to obtain sufficient quantity for sequencing and transformation into several segregants.

To ensure that no mutations were introduced into CYR1 in pJG118 and pJG121 during plasmid construction, CYR1 was sequenced from the genomic DNA isolated from segregants s4 and s33, and plasmids pJG118 and pJG121 (Genbank accession numbers for CYR1-S and CYR1-C alleles are respectively JN974464 and JN974465). Sanger sequencing was performed by the DNA Sequencing Facility at Duke University on a 3730xI DNA Analyzer (Applied Biosystems).

Plasmids pJG117, pJG118, and pJG121 were each separately transformed (Gieth and Schiestl, 2007) into segregants s2, s4, s6, s9, and s33, and selected for by growth on YPD+G418+Dox. Each transformant was grown overnight in liquid YPD+G418+Dox, cultures were washed twice with water, resuspended in water, aliquoted into five or six wells of a 96-well plate, and transferred by 96 Solid Pin Multi-Blot Replicator (V&P Scientific #VP408FP6) to YPD+G418+Dox and YEPLD+G418+Dox agar omnitrays (Nunc #242811), with doxycycline concentrations of 0, 0.01, 0.05, 1, 5, and 10 µg/mL. Omnitrays were imaged on day seven by flatbed scanner (Epson Expression 10000XL) and images were scored by visual inspection.
Complex Morphology Induction by Exogenous cAMP

YPD+cAMP and YPD+water (control) plates were inoculated with six different segregants. Triplicate colonies of each segregant were applied at 1, 2, and 3 cm from the center of the plate using a sterile wooden toothpick and incubated at 30 °C. Whole plates were imaged using a flatbed scanner (Epson Expression 10000XL) and individual colonies were imaged using a microscope (Leica MZ16) mounted digital camera (Leica DFC480).

Polystyrene Adhesion

The assay of biofilm character by quantitating polystyrene adhesion was conducted as previously described (Reynolds and Fink, 2001) with the following changes. Yeast was grown overnight (to saturation) in a 96-well plate, in 200 µL SC at 30 °C, washed twice and resuspended in 200 µL sterile H₂O. 50 µL of the cell suspension was transferred to a flat-bottom polystyrene plates containing 50 µL of 2x 0.1% Dextrose SC. The cells were incubated for 3 hours at 30 °C. Absorbance was quantitated at 590 nm (the absorbance maximum of crystal violet). A₅₉₀ values shown for each segregant are the mean of three replicate assays.

Data Availability

Datasets from this work have been archived at Dryad <http://datadryad.org/>, and can be accessed through the Digital Object Identifier <doi:10.5061/dryad.mn71g>. These datasets include results from the BSA-seq in the form of allele counts (number of occurrences of an allele in each sequencing read) for each heterozygous site found in YJM311 from the YJM311 (parental), simple, intermediate, and complex
sequencing runs, and the G-statistic values at each of these sites for the pairwise comparisons between the simple, intermediate, and complex pools. For each segregant in the assay subset data is provided on CYR1 and FLO11 genotype and expression levels, colony morphology quantitation, cAMP concentration, and polystyrene adhesion.

RESULTS

Candidate QTLs For Colony Biofilm Architecture

With our BSA-seq approach, we identified thirteen QTLs genetically linked to the variation in colony biofilm architecture (Figure 2). The top candidate genes within these QTLs are: PRP42, PPM1, FLO8, GCN1, SOL3, FLO11, YAK1, CYR1, RGT1, MSN2, HOT1, SFL1, and SKS1. By comparing the peaks inferred from the “simple versus intermediate”, “intermediate versus complex” and “simple versus complex” BSA analyses (see Methods) we were able to conclude that the very broad peak on chromosome XIII is due to overlapping signals from a major peak adjacent to HOT1, and a secondary peak centered over MSN2. (Figure 2). Six of these candidate genes (CYR1, FLO11, FLO8, SFL1, YAK1, and MSN2) cluster around the cAMP-PKA pathway, suggesting that variation in this network contributes to differential responses under distinct environmental conditions.

We chose to more closely study the effects of variation at CYR1 and FLO11, since they produced particularly strong signals in the BSA-seq, and they are respectively, a major regulator and key effector, of the cAMP-PKA pathway. We genotyped a subset of simple and complex segregants (hereafter referred to as the “assay subset”) at the CYR1 and FLO11 loci. These results showed that a complex
allele is not required at either locus to produce complex biofilm phenotypes, confirming previous evidence that biofilm formation is not a simple Mendelian trait.

**Functional Characterization of Allelic Effects at **\textit{CYR1}\**

Between the two YJM311 \textit{CYR1} alleles, we identified a combined total of 33 distinct SNPs relative to the S288c reference sequence; sixteen of these are predicted to be non-synonymous (File S3). Both alleles also share a 24 bp insertion relative to S288c. This variation and the previously known role of the cAMP–PKA pathway in colony morphology (Granek and Magwene, 2010; Halme \textit{et al.}, 2004) led us to investigate the mechanism by which \textit{CYR1} variation affects colony biofilm architecture.

Of the 70 segregants genotyped at \textit{CYR1}, 22 had the complex allele (\textit{CYR1-C}), and all were complex segregants, indicating that \textit{CYR1-C} may be sufficient for producing the complex phenotype. To test this, and determine whether cis-regulatory or coding sequence variation in \textit{CYR1} might be responsible for the morphology variation linked to the locus, we overexpressed either \textit{CYR1-S}, \textit{CYR1-C}, or \textit{lacZ} (negative control) from plasmids in several simple segregants. We found that overexpression of \textit{CYR1-S} induces complex morphology in some simple segregants, demonstrating that expression level variation is sufficient to explain the phenotypic variation linked to \textit{CYR1}. However overexpression of \textit{CYR1-C} induces morphology that is consistently more complex than that induced by \textit{CYR1-S}, indicating that coding sequence variation in \textit{CYR1} is also playing a role (Figure 3). Individual segregants respond differently to the \textit{CYR1-C} and \textit{CYR1-S} alleles and to different levels of \textit{CYR1} induction, further demonstrating that the colony biofilm phenotype results from epistatic interactions of variation acting at several loci (Figures S4 and
We measured the expression level of **CYR1** in the assay subset and found that **CYR1** expression was indeed higher in the complex segregants (Figure S5A, Mann-Whitney $p=6.148 \times 10^{-7}$). A main effect ANOVA fit to the **CYR1** expression data showed that the **CYR1** genotype explains only 7.4% of the expression variance ($p=0.04$), suggesting that variation at several of the other QTLs, *in trans* to **CYR1**, makes important contributions to its regulation.

From our genotyping and overexpression results, the functional effect of the **CYR1-C** allele was unclear. Previous work showed that biofilm formation is increased by deletion of a gene that inhibits cAMP production (**IRA2**) and by deletion of a gene that inhibits the developmental responses to cAMP (**TPK3**). Similarly, complex morphology is decreased by deletion of a gene that promotes cAMP production (**RAS2**) and by deletion of a gene that promotes developmental responses to cAMP (**TPK2**) (Granek and Magwene, 2010). We therefore hypothesized that the **CYR1-C** allele increases activity of the cAMP–PKA pathway. We tested this by measuring the cAMP concentration in colonies of each segregant in the assay subset, and found that biofilm forming segregants have significantly higher cAMP than non-biofilm forming segregants (Figure 4, Mann-Whitney $p=0.011$).

It has been demonstrated that exogenous cAMP can stimulate pseudohyphal growth, even bypassing nutrient conditions and genetic defects that otherwise suppress this developmental behavior (Lorenz *et al.*, 2000; Lorenz and Heitman, 1997). Since our results strongly indicate that increased cAMP–PKA pathway activity drives biofilm formation, we predicted that exogenous cAMP might also stimulate the formation of colony biofilms. We grew several simple segregants on media supplemented with cAMP (YPD+cAMP). The segregants tested carry the **CYR1-S** allele, but an assortment of simple and complex alleles at **FLO11**. Similar
to the pseudohyphal experiments of Lorenz et al. (Lorenz et al., 2000; Lorenz and Heitman, 1997), we found that exogenous cAMP does induce an increase in colony complexity, but as with our CYR1 overexpression experiments, the results reflect the complex genetic interactions underlying biofilm formation: when grown on YPD+cAMP, some segregants form robust colony biofilms, some moderately complex structures, and some only formed simple, non-biofilm colonies (Figure S6).

**FLO11 Expression, Polystyrene Adhesion, and Colony Biofilm Formation Are Correlated**

The adhesion protein *FLO11* and the cAMP–PKA pathway that regulates it are involved in a range of responses to stress, including formation of biofilms (Reynolds and Fink, 2001; Yi et al., 2011), filamentous and invasive growth (Lambrechts et al., 1996; Lo and Dranginis, 1998), and flocculation (Guo et al., 2000). Differential *FLO11* expression has previously been shown to contribute to variation in biofilm traits in yeast (Reynolds, 2006; Zara et al., 2009). *FLO11* expression is regulated by a complex network of interactions, which is summarized in Figure 5 (Pan and Heitman, 2002; Zhang et al., 2001; Conlan and Tzamarias, 2001). Our BSA-seq experiment identified a remarkable number of genes known to be components of this network: *CYR1, YAK1, FLO8, SFL1*, and *MSN2* (nodes highlighted in red in Fig. 5). Therefore we hypothesized that variation in this network results in differential expression of *FLO11*, which in turn contributes to the heterogeneity in biofilm architecture. We quantitated *FLO11* mRNA in the assay subset and found this to be true: *FLO11* expression is significantly higher in the complex segregants (Figure S5B; Mann-Whitney p=2.138 × 10⁻¹⁰).

We tested the segregants in the assay subset for their adherence to polystyrene,
a common measure of biofilm character (Reynolds and Fink, 2001). The genetic diversity among segregants leads to wide variation in adhesion, but we found that complex segregants were significantly more adherent than simple segregants (Figure S2; Mann-Whitney $p=1.737 \times 10^{-13}$), and that adhesion also correlates with $FLO11$ expression level (Figure S7). Since $FLO11$ is required for formation of colony biofilms (Granek and Magwene, 2010), we infer that cell-cell and cell-surface adhesion is essential, and hypothesize that variation in adhesion is one component of the architectural heterogeneity. It is therefore not surprising to find different levels of adhesion among complex segregants, and that simple segregants are poorly adherent.

**Confirmation of Candidate Gene Transcription Factors**

Three transcription factors in the cAMP-PKA network – $FLO8$, $SFL1$, and $MSN2$ – were found linked to colony biofilm formation in our QTL mapping analysis. Based on the topology of the cAMP-PKA pathway we predicted that a $\Delta flo8$ mutant would exhibit a decreased ability to form colony biofilms, while a $\Delta sfl1$ mutant should show increased biofilm complexity. Nuclear localization of $MSN2$ is inhibited by high levels of PKA activity but $MSN2$ is required for maximal $FLO11$ expression (Malcher et al., 2011). We therefore predicted a $\Delta msn2$ mutant might exhibit some loss of biofilm morphology. To test these predictions, we assayed the colony morphology of strains knocked-out for these genes (Figure S8). For this assay we used the $\Sigma 1278b$ strain background (Lorenz and Heitman, 1997), which has the convenient feature that haploids form colony biofilms, and diploids do not (Granek and Magwene, 2010), allowing us to assay for both increases and decreases in biofilm formation. As predicted, the haploid $\Delta flo8$ has a total loss of
biofilm formation, while the haploid \( \Delta msn2 \) has a substantial, but not complete loss. Also as predicted, the diploid \( \Delta sfl1/\Delta sfl1 \) shows an increase in colony complexity relative to the lack of biofilm formation in the wild-type diploid, and even the haploid \( \Delta sfl1 \) exhibits a stronger biofilm architecture relative to the wild-type haploid.

### Statistical Models Predict Strong \textit{trans} Effects on Network Function and Biofilm Traits

We fit a series of statistical models to determine the association between genotype and various measures of gene function or biofilm traits. We modeled each trait of interest using a main effect ANOVA with binary predictor variables representing the genotypes of individual segregants at each of the two major loci – \textit{CYR1} and \textit{FLO11}.

#### Colony cAMP Levels

The \textit{CYR1} genotype explains some of the variance in cAMP concentration (8.2%, \( p=0.02 \)); the \textit{FLO11} genotype has no effect (\( p=0.71 \)), as expected since it is functionally downstream of cAMP.

#### \textit{FLO11} Expression

The \textit{FLO11} allele affects its own expression (\( p=0.04 \)), but only explains 4.6% of the variance. The \textit{CYR1} allele explains a large portion of the variance in \textit{FLO11} expression (38.5%; \( p=1.47 \times 10^{-7} \)). This suggests that \textit{trans}-regulatory variation is the major driver of \textit{FLO11} expression differences between the segregants.
Plastic Adhesion

More than 60% of the adhesion variance among the segregants was explained by the alleles of \textit{CYR1} and \textit{FLO11}, (respectively 54.8\% and 7.8\%; p-values of $3.4 \times 10^{-44}$ and $6.4 \times 10^{-11}$), with most the variance attributable to \textit{CYR1}.

\section*{DISCUSSION}

We have identified thirteen candidate genes linked to variation in colony biofilm architecture, most of which were not previously implicated in the regulation of fungal biofilms. Nearly half of these genes (red in Figure 5) are members of the cAMP–PKA pathway, a single, nutrient sensitive gene network that is critical for regulating developmental responses. This network includes \textit{CYR1}, the key enzyme of the cAMP–PKA pathway; \textit{YAK1}, a kinase that is both a target of and acts in parallel to the cAMP–PKA pathway; transcription factors (\textit{FLO8}, \textit{SFL1}, and \textit{MSN2}) that are targets of both PKA and Yak1p; and \textit{FLO11}, one of the ultimate effectors of this network. This concentration of biofilm related genetic variation is particularly interesting because very few other cases have been identified (Gerke \textit{et al.}, 2009) where functional variation is clustered in a complex genetic network. While additional work will be necessary to determine the specific effects of variation at these loci on network function, we have teased apart one of the molecular mechanisms of the variation: differences in cAMP signaling.

\section*{Functional effects of \textit{CYR1} variation}

The results of the overexpression experiments indicate two mechanisms by which genetic variation at the \textit{CYR1} locus can affect biofilm formation. The first is
altered protein function, the second is differential CYR1 expression.

In the CYR1 overexpression assays the plasmids carry only the coding sequence of either CYR1-S or CYR1-C, so the difference in effects produced by the plasmids must be due to functional differences between the encoded proteins. Most of the amino acid substitutions we predict from the DNA sequence of the CYR1-S and CYR1-C alleles are novel in S. cerevisiae (File S3). The non-synonymous SNPs specific to the CYR1-C allele occur in all domains of the protein: N-terminal, Ras-binding leucine-rich repeats, catalytic domain, and Srv2p-binding region (Field et al., 1990; Kataoka et al., 1985; Nishida et al., 1998). Interestingly several of the predicted CYR1-C substitutions lead to residues that are shared with the distantly related pathogenic yeasts Candida albicans and Ashbya gossypii. All of the non-synonymous SNPs in CYR1-S (both unique to CYR1-S and shared with CYR1-C) fall in the Ras-binding domain. The large insertion shared by CYR1-C and CYR1-S is in the N-terminal domain. No function has been identified for the N-terminal and spacer domains (Yu et al., 1999), but Pfam identifies the spacer as having a protein phosphatase 2C like motif (Das et al., 1996; Finn et al., 2010). The increased cAMP–PKA pathway activity we observed in complex segregants could result from either catalytic or regulatory changes within the protein, or a combination of the two. This distribution of SNPs precludes us from identifying, based on sequence alone, which SNP is responsible for the observed phenotypic differences, and what function of Cyr1p is altered, so future experimental work will be necessary to make this determination.
Variation downstream of CYR1

Downstream of Cyr1p we identified several candidate genes that are regulated directly or indirectly by the cAMP–PKA pathway: phosphorylation by PKA directly regulates Yak1p, Flo8p and Sfl1p; Msn2p is phosphorylated and regulated by both PKA and Yak1p; *FLO11* expression is regulated by Flo8p and Sfl1p.

The kinase Yak1p is part of a cascade that acts in parallel to the cAMP–PKA pathway. It directly antagonizes the cAMP–PKA pathway through PKA: phosphorylation of the PKA regulatory subunit Bcy1p by Yak1p inhibits PKA activation (Griffioen *et al*., 2001). Yak1p is itself inactivated through phosphorylation by Tpk1p, one of the catalytic subunits of PKA (Malcher *et al*., 2011). *YAK1* plays a central role in repressing growth under glucose starvation, and promoting responses to a number of stresses including heat and acidity (Zaman *et al*., 2008; Malcher *et al*., 2011). This kinase controls stress responses though its regulation of a number of transcription factors including Hsf1p, Msn2p, Sok2p, and Phd1p (Malcher *et al*., 2011). One aspect of this stress response is the activation of surface adhesion, extracellular matrix production (Karunanithi *et al*., 2010), flocculation (Guo *et al*., 2000), filamentous and invasive growth (Lambrechts *et al*., 1996; Lo and Dranginis, 1998) as well as formation of biofilms (Reynolds and Fink, 2001), such as complex colonies (Halme *et al*., 2004; Granek and Magwene, 2010) by upregulating expression of the flocculin *FLO11* (Zhang *et al*., 2001).

The zinc-finger transcription factor Msn2p is an important regulator of general stress responses in *S. cerevisiae*. Msn2p is directly phosphorylated by PKA (Görner *et al*., 2002) and by Yak1p (Lee *et al*., 2011). Under stress or glucose limitation Msn2p and the related regulator Msn4p are localized to the nucleus (Görner *et al*., 1998). The nuclear accumulation of Msn2p is inversely related to cAMP-PKA
activity, with high-levels of cAMP-PKA activity favoring a cytoplasmic distribution. However, Msn2p is required for maximal *FLO11* expression (Malcher *et al.*, 2011) suggesting that some intermediate level of Msn2p activity may be required for biofilm formation.

Sfl1p and Flo8p regulate filamentous growth, invasive growth, and flocculation, in part through their antagonistic regulation of *FLO11* (Liu *et al.*, 1996). These transcription factors compete for a common binding site in the *FLO11* promoter, and phosphorylation by PKA has opposite effects on them: phosphorylation inhibits promoter binding by Sfl1p, and activates promoter binding by Flo8p (Pan and Heitman, 2002).

The regulation of *FLO11* expression and function is complex. Repeats within its coding sequence frequently undergo expansion and deletion that alter adhesive properties (Zara *et al.*, 2009). It is also subject to epigenetic regulation by the competitive expression of two non-coding RNAs (Bumgarner *et al.*, 2009). The *FLO11* promoter is unusually large for a *S. cerevisiae* gene and segregating variation in this promoter has been shown to affect biofilm formation in “flor yeasts” (Zara *et al.*, 2009; Fidalgo *et al.*, 2006). However, to our knowledge, we are the first to identify extensive variation in trans to *FLO11* that affects its regulation. The strong correlations of *FLO11* expression levels with colony morphology and polystyrene adhesion are further indications that this network variation is functional with respect to formation of colony biofilms.

**A Model Relating Variation in cAMP Signaling to Colony Biofilms**

Based on our findings and current understanding of the cAMP-PKA pathway, we predict that the allelic variation we identified across this network has the fol-
lowing effects that are linked to colony biofilm architecture: (1) increased adeny-
late cyclase expression and/or activity; (2) increased intracellular levels of cAMP; (3) increased PKA activity; (4) increased activity of transcriptional activator Flo8p; (5) decreased activity of transcriptional repressor Sfl1p; (6) increased *FLO11* expression; (7) increased cell-cell and cell-substrate adhesion. Our analysis of the
effect of *CYR1* overexpression, and our comparison of cAMP concentrations in
simple versus complex colonies support predictions 1-3. Similarly, our analysis of
the transcription factor knockout mutants, Δflo8 and Δsfl1, supports predictions 4
and 5. Finally, the analysis of *FLO11* expression and the assay of plastic adherence
support predictions 6 and 7.

Our predictions regarding allelic variation at *YAK1* and *MSN2* are somewhat
more complicated. Malcher (Malcher *et al.*, 2011) proposed a model in which
Yak1p is inhibited under abundant glucose and high-PKA conditions, but can pro-
mote *FLO11* expression and cellular adhesion under stressed and glucose-limiting
conditions. Similarly, high PKA activity is predicted to inhibit nuclear accumu-
luation of Msn2p, but *MSN2* is required for maximal *FLO11* expression (Malcher
*et al.*, 2011). Some degree of Msn2p activity would also help increase general stress
resistance and promote continued growth under stressful conditions. We there-
fore predict that *YAK1* and *MSN2* are particularly likely to show complex epistatic
interactions with variation at other loci across the cAMP-PKA pathway. The very
weak biofilm formed by the Δ*msn2* mutant, (a decrease relative to the WT haploid),
support this (Figure S8.)
Other Candidate Loci for Biofilm Variation

It is clear from our BSA-seq results that variation in the mechanisms for sensing and responding to nutrients is central to the diversity we observe in biofilms architecture. In addition to the variation in the cAMP-PKA-FLO11 network, we find that most of the remaining candidate genes are involved in detecting and utilizing nutrients. RGT1 and SOL3 both play roles in glucose utilization (Ozcan and Johnston, 1995; Stanford et al., 2004); RGT1 was previously show to be necessary for complex morphology in the Σ1278b background (Granek and Magwene, 2010). PPM1 and GCN1 both help regulate the response to nitrogen (Wu et al., 2000; Garcia-Barrio et al., 2000). Both are linked to the TOR signaling pathway, providing further indication that the TOR pathway plays a role in regulating colony biofilm formation, in addition to the link through RIM15, which was previously demonstrated (Granek and Magwene, 2010).

Interactions Between cAMP Signaling, Glucose Starvation, and Biofilm Formation

We have previously shown that limitation for fermentable carbon sources is a trigger for the formation of colony biofilms (Granek and Magwene, 2010). Since cAMP production is increased in response to abundant fermentable carbon, it is somewhat counterintuitive that low glucose and high cAMP-PKA signaling are both triggers. We propose, and our data supports, the hypothesis that strains that produce biofilms are precisely those backgrounds that are able to maintain relatively high levels of cAMP-PKA signaling in the face of glucose starvation. Since the cAMP-PKA pathway is primarily a ‘pro-growth’ pathway in *S. cerevisiae*, it is
tempting to speculate that the ruffles and folds characteristic of colony biofilms are due to spatial heterogeneity in cell division rates across the colony, enabled by this upregulation of cAMP signaling. Palková and colleagues have recently presented evidence of such spatial heterogeneity of division rates and other metabolic properties within yeast colonies (Stovíček et al., 2010; Cáp et al., 2012).

**Heterozygous Diploid Bulk Segregant Analysis**

Recent studies have demonstrated that the genomes of many clinical and industrial isolates of *S. cerevisiae* have high levels of heterozygosity (Magwene et al., 2011a; Argueso et al., 2009; Akao et al., 2011). This heterozygosity underlies a huge range of phenotypic variation. The novel “heterozygous diploid” BSA strategy employed here can be readily used to exploit this heterozygosity in order to map QTLs for traits relevant in medical and industrial settings. Importantly, the variability generated from such strains is expected to replicate the range produced when meiosis and selfing occur in a natural environment. This is a key difference from the standard BSA method, which depends on crossing divergent genetic backgrounds which may have never come into contact outside the laboratory, and therefore may produce progeny with allelic combinations and phenotypes unlikely to occur in nature.

**The Utility of Multiple Bulks in BSA Studies**

Another novel feature of our BSA approach is the use of multiple bulk comparisons. Most BSA studies involve single pairwise comparisons, such as between a high and low bulk or between a selected and control bulk. In the present study we employed a three bulk strategy, with low (simple), intermediate, and high (com-
plex) pools of segregants. Using this approach we were able to show that what appeared to be a single large peak on chromosome XIII in the simple-versus-complex bulk comparison was instead two distinct peaks, as revealed by the simple-versus-intermediate and intermediate-versus-complex comparisons (Figure 2). Comparisons involving the intermediate bulk also suggest more subtle relationships between genotype and phenotype. For example, the MSN2 peak is almost entirely due to differences in allele frequency in the transition from simple to intermediate morphology. Similarly, the CYR1 peak appears in the transition from intermediate to complex morphology. Ongoing studies in our laboratory (Magwene and Zeyl, unpublished results) suggest these types of findings may be a common feature of complex trait variation in *S. cerevisiae*, and we propose that this is likely to be true in general. Therefore, we recommend that future BSA-seq studies incorporate multiple bulk comparisons whenever feasible.

**Conclusions**

The variety of colony phenotypes segregating in YJM311, the number of genes identified as contributing to colony biofilm architecture, and the differential response of segregants to CYR1 overexpression and exogenous cAMP all make clear that formation of colony biofilms is the product of multiple processes. At the same time, the fact that many of the candidate genes we identified are part of an integrated signaling cascade suggests that these processes converge on a few major integration points through which genetic variation can act. If this is indeed the case then one would predict that mapping studies employing backgrounds distantly related to YJM311 are likely to identify causal variants at different genes, but that most of these genes may converge on the same core pathways. Future studies that
exploit the wide diversity of colony phenotypes present among lineages of *S. cerevisiae* will help to shed light on the structure of variation and provide novel insight both into the genetic basis of complex traits and gene network evolution.

**ACKNOWLEDGEMENTS**

We wish to acknowledge Joseph Heitman, John McCusker, Colin Maxwell, and two anonymous reviewers for helpful comments on the manuscript, Aimee Zhang for work related to this project, and the Duke University Institute for Genome Sciences & Policy Genome Sequencing & Analysis Core Resource for Sanger and Next-Generation sequencing.
LITERATURE CITED


Goldman, D. L., B. C. Fries, S. P. Franzot, L. Montella, and A. Casadevall, 1998 Phenotypic switching in the human pathogenic fungus Cryptococcus neoformans is


Lorenz, M. C., and J. Heitman, 1997 Yeast pseudohyphal growth is regulated by GPA2, a G protein alpha homolog. EMBO J 16: 7008–18.


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


**Sherman, F., 2002 Getting started with yeast.** Meth Enzymol 350: 3–41.

St’ovíˇcek, V., L. Váchová, M. Kuthan, and Z. Palková, 2010 General factors important for the formation of structured biofilm-like yeast colonies. Fungal Genet Biol.


Figure 1: Segregants of the clinical isolate YJM311 are very heterogeneous for colony morphology. (A) YJM311 and (B) six of its random segregants demonstrate a variety of colony morphologies. (C) A representative probability density estimate of YJM311 random segregant colony morphology scores. The rug plot at bottom indicates values for individual segregants.

Figure 2: Thirteen loci show significant linkage to colony morphology variation. The $G'$-statistic is derived from the frequency of alleles at each variable site, estimated by pairwise comparison of bulks: simple versus intermediate (blue); intermediate versus complex (red); and simple versus complex (magenta). Peak labels indicate candidate genes within each QTL predicted to contribute to the observed phenotypic variation. Discontinuities in the plot indicate large regions of the genome that lack heterozygous SNPs.

Figure 3: CYR1 overexpression induces complex morphology in a genotype dependent manner. Simple colony, CYR1-S segregants S2, S4, S6, and S9, representing a variety of genotypes at the FLO11 loci were transformed with plasmids that express lacZ, CYR1-S, or CYR1-C from a tetracycline-regulatable promoter. Segregants were grown on YPD+G418.CYR1-S and CYR1-C both induce complex morphology by day seven, although CYR1-C has a stronger effect. Overexpression of lacZ (negative control) has no effect on colony morphology.

Figure 4: Complex segregants produce higher levels of cAMP.

The probability density estimates for distribution of cAMP concentration in sim-
ple (black) and complex segregants (red). Complex segregants have significantly higher cAMP than simple segregants (Mann-Whitney, p=0.011). Colonies formed by YJM311 segregants were weighed, then cAMP in each colony was quantitated by ELISA to determine the amount of cAMP per mg of cells. The rug plot at bottom shows values for individual samples.

**Figure 5: Segregating variation at multiple levels of the cAMP-PKA network contributes to colony morphology heterogeneity.** The Bulk Segregant Analysis indicates that functional variation is concentrated in the cAMP-PKA network (candidate QTGs are red), and this variation is a major contributor to biofilm heterogeneity. Most of the candidate QTGs (red) in the network are known to promote the biofilm associated behaviors adhesion and stress responses (*CYR1, FLO8, FLO11, YAK1, MSN2*), while *SFL1* suppresses adhesion. The variant alleles seem to attenuate or reverse previously identified activities. Arrows indicate activating effects, bars indicate suppressive effects. Dashed lines summarize one or more intermediate interactions that are not shown.
FIGURES
Figure 1

A

B

C

Probability Density

Colony Morphology
Figure 2
Figure 3
Figure 4

The figure shows a probability density plot of cAMP concentration (fmol of cAMP/mg of cells) for two conditions: simple and complex. The x-axis represents the cAMP concentration, ranging from 150 to 300 fmol/mg cells, while the y-axis represents the probability density, ranging from 0.000 to 0.008. The plot includes two curves, one for simple conditions (black) and one for complex conditions (red), illustrating differences in probability density across the concentration range.
Figure 5

Stress

Glucose

ATP → cAMP

Yak1

PKA

Stress Responses

Msn2

Stress Glucose

Growth Responses Adhesion

Flo8 Sfl1

Flo11

Adhesion
### Tables

**Table 1: Strains**

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>YJM311</td>
<td>aro/ARO MAT(\alpha/\alpha)</td>
<td>(McCusker et al., 1994a)</td>
</tr>
<tr>
<td>MLY40(\alpha)</td>
<td>ura3-52 MAT(\alpha)</td>
<td>(Lorenz and Heitman, 1997)</td>
</tr>
<tr>
<td>MLY41a</td>
<td>ura3-52 MAT(a)</td>
<td>(Lorenz and Heitman, 1997)</td>
</tr>
<tr>
<td>MLY61a/(\alpha)</td>
<td>ura3-52/ura3-52</td>
<td>(Lorenz and Heitman, 1997)</td>
</tr>
<tr>
<td>XPY95a</td>
<td>(\Delta)flo8::HygB ura3-52 MAT(\alpha)</td>
<td>(Pan and Heitman, 1999)</td>
</tr>
<tr>
<td>XPY95(\alpha)</td>
<td>(\Delta)flo8::HygB ura3-52 MAT(\alpha)</td>
<td>(Pan and Heitman, 1999)</td>
</tr>
<tr>
<td>XPY95a/(\alpha)</td>
<td>(\Delta)flo8::HygB/(\Delta)flo8::HygB ura3-52/ura3-52 MAT(\alpha/\alpha)</td>
<td>(Pan and Heitman, 1999)</td>
</tr>
<tr>
<td>XPY108a</td>
<td>(\Delta)sfl1::HygB ura3-52 MAT(\alpha)</td>
<td>(Pan and Heitman, 2002)</td>
</tr>
<tr>
<td>XPY108(\alpha)</td>
<td>(\Delta)sfl1::HygB ura3-52 MAT(\alpha)</td>
<td>(Pan and Heitman, 2002)</td>
</tr>
<tr>
<td>XPY108a/(\alpha)</td>
<td>(\Delta)sfl1::HygB/(\Delta)sfl1::HygB ura3-52/ura3-52 MAT(\alpha/\alpha)</td>
<td>(Pan and Heitman, 2002)</td>
</tr>
<tr>
<td>MLY226a</td>
<td>(\Delta)msn2::G418 ura3-52 MAT(\alpha)</td>
<td>(Pan and Heitman, 1999)</td>
</tr>
<tr>
<td>MLY226(\alpha)</td>
<td>(\Delta)msn2::G418 ura3-52 MAT(\alpha)</td>
<td>(Pan and Heitman, 1999)</td>
</tr>
<tr>
<td>MLY226a/(\alpha)</td>
<td>(\Delta)msn2::G418/(\Delta)msn2::G418 ura3-52/ura3-52 MAT(\alpha/\alpha)</td>
<td>(Pan and Heitman, 1999)</td>
</tr>
<tr>
<td>BY4743</td>
<td>his3A(\Delta)/his3A(\Delta), leu2A(\Delta)/leu2A(\Delta), lys2A(\Delta)/LYS2, MET15/met15A(\Delta), ura3A(\Delta)/ura3A(\Delta) MAT(\alpha/\alpha)</td>
<td>(Brachmann et al., 1998)</td>
</tr>
</tbody>
</table>

**Table 2: Plasmids**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Use</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pCM179</td>
<td>Basis for (\text{lacZ}) expression plasmid</td>
<td>(Garí et al., 1997)</td>
</tr>
<tr>
<td>pCM190</td>
<td>Basis for (\text{CYR1}) expression plasmids</td>
<td>(Garí et al., 1997)</td>
</tr>
<tr>
<td>pJG117</td>
<td>Tet-regulated (\text{lacZ}) expression plasmid</td>
<td>This study</td>
</tr>
<tr>
<td>pJG118</td>
<td>Tet-regulated (\text{CYR1-S}) expression plasmid</td>
<td>This study</td>
</tr>
<tr>
<td>pJG121</td>
<td>Tet-regulated (\text{CYR1-C}) expression plasmid</td>
<td>This study</td>
</tr>
</tbody>
</table>