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

Joshua A. Granek, Debra Murray, Ömür Kayrıkçı, and Paul M. Magwene¹

Department of Biology and Center for Systems Biology, Duke University, Durham, North Carolina 27708

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 13 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 overexpression, 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. They are 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 mechanistic model that relates genetic variation to gene network function and phenotypic outcomes.

BIOFILMS are complex, surface-adherent communities of microbes. They are ubiquitous in nature and in the human environment (Donlan and Costerton 2002; López *et al.* 2010). 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* 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 nonbiofilm colonies (also described as simple or smooth colonies) by intricate, organized, and strain-specific architecture (Supporting Information, Figure S1) (Palková and Váchová 2006; Granek and Magwene 2010; 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 *Saccharomyces cerevisiae* in the propensity to form biofilms, the architectures produced, and the effect of environmental signals on biofilm phenotypes (*i.e.*, genotype-by-environment

Copyright © 2013 by the Genetics Society of America

doi: 10.1534/genetics.112.142067

Manuscript received May 23, 2012; accepted for publication November 1, 2012

Supporting information is available online at <http://www.genetics.org/lookup/suppl/>

doi:10.1534/genetics.112.142067/-/DC1.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. JN974464 and JN974465.

¹Corresponding author: Department of Biology, Duke University, P.O. Box 90338, Durham, NC 27708. E-mail: paul.magwene@duke.edu

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 (Clemons *et al.* 1994; McCusker *et al.* 1994b). 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 nonbiofilm 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) (Michelmore *et al.* 1991; Brauer *et al.* 2006; Schneeberger *et al.* 2009; Ehrenreich *et al.* 2010; 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 (>40,000 heterozygous SNPs, representing >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 a 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

Figure 1 Segregants of the clinical isolate YJM311 are very heterogeneous for colony morphology. (A and B) YJM311(A) and six of its random segregants (B) 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.

apart multiple overlapping quantitative trait loci (QTL) that would otherwise have been impossible to disambiguate had we used only the extreme bulks.

Using BSA-seq we mapped 13 colony biofilm QTL and identified a candidate gene associated with each QTL peak. Of these 13 genes, 6 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*, encodes 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 regulated directly 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 provide novel insights into the role of network variation on complex traits. Because the cAMP–PKA pathway is a 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 were made with 200 mg/liter Geneticin. Tetracycline-regulating “Dox” media contained 10 mg/liter doxycycline, except where otherwise indicated. YPD+cAMP and YPD+water media were prepared by spreading 667 μ l of 30 mM cAMP (A9501-1G; Sigma Aldrich, St. Louis), yielding a final concentration of 1 mM cAMP, or 667 μ l 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 35 segregants randomly selected from the simple pool and 35 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) with the following changes. Cells grown to saturation in YPD at 30° overnight were washed, resuspended in sterile water, plated on sporulation media, and incubated at room temperature for 9 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 resuspended 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°. Once grown, the morphology of colonies was judged by visual inspection for assignment to a pool, and they were picked, suspended in 100 μ l YPD in a 96-well plate, and incubated overnight at 30°. Segregants were replicated from the 96-well plate to a YEPLD agar omni-tray (Nunc no. 242811), using a 96 Solid Pin Multi-Blot Replicator (V&P Scientific no. 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° for 2 days. Individual cultures from each 96-well plate were aspirated and pooled in a preweighed 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 and pelleted again, the supernatant was again removed, the conical tube was weighed to determine the total cell mass, and the final pellet was frozen at –80°. DNA was extracted from the pools, using anion-exchange gravity flow columns (Genomic-Tip; QIAGEN, Valencia, CA). To construct the genomic libraries, we used ~5 μ g of material from each pool and followed Illumina Sample Preparation Kit instructions. Libraries were quantified using Experion electrophoresis (Bio-Rad, Hercules, CA) and a Qubit fluorometer (Invitrogen, Carlsbad, CA) 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 numbers of reads for each pool were 30.3 M for simple, 14.5 M for intermediate, and 31.7 M for complex. Short reads were mapped to the standard *S. cerevisiae* reference genome as

Table 1 Strains

Strain	Genotype	Reference
YJM311	<i>aro/ARO MATa</i> α	McCusker <i>et al.</i> (1994a)
MLY40 α	<i>ura3-52 MATa</i> α	Lorenz and Heitman (1997)
MLY41a	<i>ura3-52 MATa</i>	Lorenz and Heitman (1997)
MLY61a/ α	<i>ura3-52/ura3-52</i>	Lorenz and Heitman (1997)
XPY95a	Δ <i>flo8::HygB ura3-52 MATa</i>	Pan and Heitman (1999)
XPY95 α	Δ <i>flo8::HygB ura3-52 MATa</i> α	Pan and Heitman (1999)
XPY95a/ α	Δ <i>flo8::HygB/Δflo8::HygB ura3-52/ura3-52 MATa</i> α	Pan and Heitman (1999)
XPY108a	Δ <i>sfl1::HygB ura3-52 MATa</i>	Pan and Heitman (2002)
XPY108 α	Δ <i>sfl1::HygB ura3-52 MATa</i> α	Pan and Heitman (2002)
XPY108a/ α	Δ <i>sfl1::HygB/Δsfl1::HygB ura3-52/ura3-52 MATa</i> α	Pan and Heitman (2002)
MLY226a	Δ <i>msn2::G418 ura3-52 MATa</i>	Pan and Heitman (1999)
MLY226 α	Δ <i>msn2::G418 ura3-52 MATa</i> α	Pan and Heitman (1999)
MLY226a/ α	Δ <i>msn2::G418/Δmsn2::G418 ura3-52/ura3-52 MATa</i> α	Pan and Heitman (1999)
BY4743	<i>his3Δ1/his3Δ1, leu2Δ0/leu2Δ0, lys2Δ0/LYS2, MET15/met15Δ0, ura3Δ0/ura3Δ0 MATa</i> α	Brachmann <i>et al.</i> (1998)

described above. The average depths of coverage for each pool, based on mapped reads, were 120 $\times$ for simple, 55 $\times$ for intermediate, and 126 $\times$ for complex.

We sequenced the heterozygous diploid strain YJM311, using 50-bp reads on an Illumina GAII Genome Analyzer at the Duke University DNA Sequencing Facility. Sequencing yielded $\sim$ 15 M 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.* (2011a), which was based on 36-bp 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 were 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, and 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, postsequencing alignment of reads, etc.). Magwene *et al.* (2011b) describe a statistical and analytical framework that accounts for these two sources of var-

iation 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 10 SNPs, spanning at least 10 kb, where the *G'* statistic exceeded the false discovery rate threshold. Peaks were identified as contiguous subregions such that $G'_i \geq 0.90G'$, where G'_{max} is the site that has the largest *G'* within the QTL region. Details of statistically significant peaks are provided in File S2. For each QTL peak we identified candidate genes based on criteria such as the number of variable sites, the number of predicted nonsynonymous 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° and imaged by flatbed scanner (Epson Expression 10000XL). Images were then scored by visual inspection; scores were assigned on a scale from 1 to 5 to one decimal place, with 1 indicating completely simple, nonbiofilm phenotypes and 5 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

Table 2 Plasmids

Plasmid	Use	Reference
pCM179	Basis for <i>lacZ</i> expression plasmid	Garí <i>et al.</i> (1997)
pCM190	Basis for <i>CYR1</i> expression plasmids	Garí <i>et al.</i> (1997)
pJG117	Tet-regulated <i>lacZ</i> expression plasmid	This study
pJG118	Tet-regulated <i>CYR1-S</i> expression plasmid	This study
pJG121	Tet-regulated <i>CYR1-C</i> expression plasmid	This study

Figure S3 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 3 days at 30°. Individual colonies were then scraped from plates, suspended in 500 µl 1 M *n*-butanol-saturated formic acid to fix, and 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 nonacetylation 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 3 days at 30°. Individual colonies were then scraped from plates and suspended in ice-cold water. Cells were pelleted by centrifugation, supernatant was removed, and pellets were flash frozen in liquid nitrogen and stored at -80°. 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 high-resolution melt (HRM) analysis. HRM primers were designed such that they had melting temperatures of 56° and that pairs had similar melting temperatures and would amplify 100- to 150-bp 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 File S1. The SNPs assayed are detailed in Table S1; coordinates in Table S1 are based on *Saccharomyces* Genome Database release r64 (release date: February 4, 2011) (Cherry *et al.* 1997). Sanger sequencing was performed by the Duke University Institute for Genome Sciences and Policy Genome Sequencing and Analysis Core Resource on a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA).

Genomic DNA templates for HRM were isolated from segregants, using the DNeasy Blood and 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 Qiagen AE Buffer and then quantified with an ND-1000 Spectrophotometer (NanoDrop) 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 10-µl 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° increments in the initial range of 65°–95°, with a 2-sec 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 affect the overall HRM analysis, we removed any outliers prior to data analysis. This includes samples that amplified late (*C_t* values >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 runs (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” (S).

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 intratetrad mating. It is possible that the heterozygotes arose from diploids that survived the random sporulation procedure. This is unlikely, since no colonies grew in 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 using 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 FidelityTaq (United States Biochemical, Cleveland, OH), using primers pcm_kan_up and pcm190_kan_dn, which have 50-bp 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 and 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 50-bp 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. Twenty-five-microliter PCR reactions consisted of Phusion Buffer GC, 200 μ M each dNTP, 0.5 μ M each primer, 100–150 ng genomic DNA, 5% DMSO, 2.1 mM MgCl₂ (total), and 0.75 unit Phusion Polymerase. The cycling program was 2 min at 98°; 25 cycles of 15 sec at 98°, 30 sec at 59°, and 190 sec at 72°; with a final extension of 10 min at 72°; 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 cotransformed into strain BY4743, using the lithium acetate method (Gietz and Schiestl 2007). pJG117 was generated by cotransforming pCM179 with the KanMX4 PCR product. pJG118 and pJG121 were assembled by cotransforming 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*, to avoid selecting against functional plasmids, which might have a higher fitness cost than nonfunctional plasmids. Plasmids were rescued from yeast by the QIAprep Spin Miniprep Kit (QIAGEN), following the manufacturer's protocol for yeast; transformed into *Escherichia coli* (ElectroSHOX Competent Cells; Bionline); and purified by the QIAprep Spin Miniprep Kit (QIAGEN) to obtain a 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 3730xl DNA Analyzer (Applied Biosystems).

Plasmids pJG117, pJG118, and pJG121 were each separately transformed (Gietz 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, and cultures were washed twice with water, resuspended in water, aliquoted into 5 or 6 wells of a 96-well plate, and transferred by a 96 Solid Pin Multi-Blot Replicator (V&P Scientific no. VP408FP6) to YPD+G418+Dox and YEPLD+G418+Dox agar omnitrays (Nunc no. 242811), with doxycycline concentrations of 0, 0.01, 0.05, 1, 5, and 10 μ g/ml. Omnitrays were imaged on day 7 by a flat-bed 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°. Whole plates were imaged using a flatbed scanner (Epson Expression 10000XL) and individual colonies were imaged using a digital camera (Leica DFC480) mounted on a microscope (Leica MZ16).

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 μ l SC at 30°, washed twice, and resuspended in 200 μ l sterile H₂O. Fifty microliters of the cell suspension was transferred to a flat-bottom polystyrene plate containing 50 μ l of 2 $\times$ 0.1% Dextrose SC. The cells were incubated for 3 hr at 30°. Absorbance was quantitated at 590 nm (the absorbance maximum of crystal violet). *A*₅₉₀ values shown for each segregant are the mean of three replicate assays.

Data availability

Data sets 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 data sets 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

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 vs. intermediate (blue), intermediate vs. complex (red), and simple vs. 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.

these sites for the pairwise comparisons between the simple, intermediate, and complex pools. For each segregant in the assay subset data are provided on *CYR1* and *FLO11* genotype and expression levels, colony morphology quantitation, cAMP concentration, and polystyrene adhesion.

Results

Candidate QTL for colony biofilm architecture

With our BSA-seq approach, we identified 13 QTL genetically linked to the variation in colony biofilm architecture (Figure 2). The top candidate genes within these QTL are *PRP42*, *PPM1*, *FLO8*, *GCN1*, *SOL3*, *FLO11*, *YAK1*, *CYR1*, *RGT1*, *MSN2*, *HOT1*, *SFL1*, and *SKS1*. By comparing the peaks inferred from the “simple vs. intermediate,” “intermediate vs. complex,” and “simple vs. complex” BSA analyses (see *Materials and 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 a 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 *CYR1*

Between the two YJM311 *CYR1* alleles, we identified a combined total of 33 distinct SNPs relative to the S288c

reference sequence; 16 of these are predicted to be nonsynonymous (File S3). Both alleles also share a 24-bp insertion relative to S288c. This variation and the previously known role of the cAMP–PKA in colony morphology (Halme *et al.* 2004; Granek and Magwene 2010) led us to investigate the mechanism by which *CYR1* variation affects colony biofilm architecture.

Of the 70 segregants genotyped at *CYR1*, 22 had the complex allele (*CYR1*-C), and all were complex segregants, indicating that *CYR1*-C may be sufficient for producing the complex phenotype. To test this and determine whether *cis*-regulatory or coding sequence variation in *CYR1* might be responsible for the morphology variation linked to the locus, we overexpressed *CYR1*-S, *CYR1*-C, or *lacZ* (negative control) from plasmids in several simple segregants. We found that overexpression of *CYR1*-S induces complex morphology in some simple segregants, demonstrating that expression level variation is sufficient to explain the phenotypic variation linked to *CYR1*. However, overexpression of *CYR1*-C induces morphology that is consistently more complex than that induced by *CYR1*-S, indicating that coding sequence variation in *CYR1* is also playing a role (Figure 3). Individual segregants respond differently to the *CYR1*-C and *CYR1*-S alleles and to different levels of *CYR1* induction, further demonstrating that the colony biofilm phenotype results from epistatic interactions of variation acting at several loci (Figure 3 and Figure S3).

We measured the expression level of *CYR1* in the assay subset and found that *CYR1* expression was indeed higher in the complex segregants (Figure S4A, Mann–Whitney $P = 6.148 \times 10^{-7}$). A main-effect ANOVA fitted 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 QTL, 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

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 7, although *CYR1-C* has a stronger effect. Overexpression of *lacZ* (negative control) has no effect on colony morphology.

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 and Heitman 1997; Lorenz *et al.* 2000). 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 and Heitman 1997; Lorenz *et al.* 2000), 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 form moderately complex structures, and some form only simple, nonbiofilm colonies (Figure S5).

***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,

Figure 4 Complex segregants produce higher levels of cAMP. Shown are the probability density estimates for distribution of cAMP concentration in simple (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 and then cAMP in each colony was quantitated by ELISA to determine the amount of cAMP per milligram of cells. The rug plot at bottom shows values for individual samples.

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 (Conlan and Tzamarias 2001; Zhang *et al.* 2001; Pan and Heitman 2002). 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 Figure 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 S4B; Mann–Whitney $P = 2.138 \times 10^{-10}$).

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 S6). Since *FLO11* is required for formation of colony biofilms (Granek and Magwene 2010), we infer that cell–cell adhesion and cell–surface adhesion are essential and hypothesize that variation in adhesion

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 trait genes (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 seems to attenuate or reverse previously identified activities. Arrows indicate activating effects, and bars indicate suppressive effects. Dashed lines summarize one or more intermediate interactions that are not shown.

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 S6). 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 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: *CYR1* and *FLO11*.

Colony cAMP levels

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

FLO11 expression

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

Plastic adhesion

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

Discussion

We have identified 13 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 *CYR1*, the key enzyme of the cAMP-PKA pathway; *YAK1*, a kinase that both is a target of and acts in parallel to the cAMP-PKA pathway; transcription factors (*FLO8*, *SFL1*, and *MSN2*) that are targets of both PKA and *Yak1p*; and *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 *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.

Functional effects of *CYR1* variation

The results of the overexpression experiments indicate two mechanisms by which genetic variation at the *CYR1* locus can affect biofilm formation. The first is altered protein function, and 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 nonsynonymous SNPs specific to the *CYR1-C* allele occur in all domains of the protein: N-terminal, Ras-binding leucine-rich repeats, the catalytic domain, and the *Srv2p*-binding region (Kataoka *et al.* 1985; Field *et al.* 1990; Nishida *et al.* 1998). Interestingly, several of the predicted *CYR1-C* substitutions lead to residues that are shared with the distantly related pathogenic yeasts *C. albicans* and *Ashbya gossypii*. All of the nonsynonymous 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*; and *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 through 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), and 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 noncoding 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” (Fidalgo *et al.* 2006; Zara *et al.* 2009). 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 following effects that are linked to colony biofilm architecture: (1) increased adenylate 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, and (7) increased cell–cell and

cell–substrate adhesion. Our analysis of the effect of *CYR1* overexpression and our comparison of cAMP concentrations in simple vs. complex colonies support predictions 1–3. Similarly, our analysis of the transcription factor knockout mutants, $\Delta flo8$ and $\Delta sfl1$, 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 promote *FLO11* expression and cellular adhesion under stressed and glucose-limiting conditions. Similarly, high PKA activity is predicted to inhibit nuclear accumulation 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 therefore 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 $\Delta msn2$ mutant (a decrease relative to the WT haploid) supports this (Figure S6).

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 shown to be necessary for complex morphology in the $\Sigma 1278b$ background (Granek and Magwene 2010). *PPM1* and *GCN1* both help regulate the response to nitrogen (Garcia-Barrio *et al.* 2000; Wu *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 support, 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 (Šťovíček *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 (Argueso *et al.* 2009; Akao *et al.* 2011; Magwene *et al.* 2011a). 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 to map QTL 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 that 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 a low bulk or between a selected and a control bulk. In the present study we employed a three-bulk strategy, with low (simple), intermediate, and high (complex) 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 vs. complex bulk comparison was instead two distinct peaks, as revealed by the simple vs. intermediate and intermediate vs. 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 (P. M. Magwene and C. 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 into both the genetic basis of complex traits and gene network evolution.

Acknowledgments

We 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 and Policy Genome Sequencing and Analysis Core Resource for Sanger and next-generation sequencing. This work was supported by an NIH award to P. M. M. (P50GM081883-01).

Literature Cited

- Akao, T., I. Yashiro, A. Hosoyama, H. Kitagaki, H. Horikawa *et al.*, 2011 Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* kyokai no. 7. *DNA Res.* 18: 423–434.
- Argueso, J. L., M. F. Carazzolle, P. A. Mieczkowski, F. M. Duarte, O. V. C. Netto *et al.*, 2009 Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res.* 19: 2258–2270.
- Brachmann, C. B., A. Davies, G. J. Cost, E. Caputo, J. Li *et al.*, 1998 Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14: 115–132.
- Brauer, M. J., C. M. Christianson, D. A. Pai, and M. J. Dunham, 2006 Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. *Genetics* 173: 1813–1816.
- Bumgarner, S. L., R. D. Dowell, P. Grisafi, D. K. Gifford, and G. R. Fink, 2009 Toggle involving *cis*-interfering noncoding RNAs controls variegated gene expression in yeast. *Proc. Natl. Acad. Sci. USA* 106: 18321–18326.
- Cáp, M., L. Stěpánek, K. Harant, L. Váchová, and Z. Palková, 2012 Cell differentiation within a yeast colony: metabolic and regulatory parallels with a tumor-affected organism. *Mol. Cell* 46: 436–448.
- Cherry, J. M., C. Ball, S. Weng, G. Juvik, R. Schmidt *et al.*, 1997 Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* 387: 67–73.
- Clemons, K. V., J. H. McCusker, R. W. Davis, and D. A. Stevens, 1994 Comparative pathogenesis of clinical and nonclinical isolates of *Saccharomyces cerevisiae*. *J. Infect. Dis.* 169: 859–867.
- Collart, M. A., and S. Oliviero, 2001 Preparation of yeast RNA, pp. 13.12.1–13.12.5 in *Current Protocols in Molecular Biology*, edited by F. M. Ausubel, R. Brent, R. R. Kingston, D. D. Moore, J. G. Seidman *et al.* John Wiley & Sons, New York.
- Conlan, R. S., and D. Tzamarias, 2001 Sfl1 functions via the co-repressor Ssn6-Tup1 and the cAMP-dependent protein kinase Tpk2. *J. Mol. Biol.* 309: 1007–1015.
- Das, A. K., N. R. Helps, P. T. Cohen, and D. Barford, 1996 Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* 15: 6798–6809.
- Donlan, R. M., and J. W. Costerton, 2002 Biofilms: survival mechanisms of clinically relevant microorganisms. *Clin. Microbiol. Rev.* 15: 167–193.
- Ehrenreich, I. M., N. Torabi, Y. Jia, J. Kent, S. Martis *et al.*, 2010 Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464: 1039–1042.
- Fidalgo, M., R. R. Barrales, J. I. Ibeas, and J. Jimenez, 2006 Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl. Acad. Sci. USA* 103: 11228–11233.
- Field, J., H. P. Xu, T. Michaeli, R. Ballester, P. Sass *et al.*, 1990 Mutations of the adenyl cyclase gene that block RAS function in *Saccharomyces cerevisiae*. *Science* 247: 464–467.
- Finn, R. D., J. Mistry, J. Tate, P. Coghill, A. Heger *et al.*, 2010 The Pfam protein families database. *Nucleic Acids Res.* 38: D211–D222.
- Flemming, H.-C., 2002 Biofouling in water systems—cases, causes and countermeasures. *Appl. Microbiol. Biotechnol.* 59: 629–640.
- Francis, S. H., and J. D. Corbin, 1994 Structure and function of cyclic nucleotide-dependent protein kinases. *Annu. Rev. Physiol.* 56: 237–272.
- Garcia-Barrio, M., J. Dong, S. Ufano, and A. G. Hinnebusch, 2000 Association of GCN1–GCN20 regulatory complex with the N-terminus of eIF2 α kinase GCN2 is required for GCN2 activation. *EMBO J.* 19: 1887–1899.
- Garí, E., L. Piedrafita, M. Aldea, and E. Herrero, 1997 A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* 13: 837–848.
- Geiss, G. K., R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar *et al.*, 2008 Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* 26: 317–325.
- Gerke, J., K. Lorenz, and B. Cohen, 2009 Genetic interactions between transcription factors cause natural variation in yeast. *Science* 323: 498–501.
- Gietz, R. D., and R. H. Schiestl, 2007 High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* 2: 31–34.
- Goddard, M. R., H. C. J. Godfray, and A. Burt, 2005 Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* 434: 636–640.
- 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 associated with changes in virulence and pulmonary inflammatory response in rodents. *Proc. Natl. Acad. Sci. USA* 95: 14967–14972.
- Görner, W., E. Durchschlag, M. T. Martinez-Pastor, F. Estruch, G. Ammerer *et al.*, 1998 Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. *Genes Dev.* 12: 586–597.
- Görner, W., E. Durchschlag, J. Wolf, E. L. Brown, G. Ammerer *et al.*, 2002 Acute glucose starvation activates the nuclear localization signal of a stress-specific yeast transcription factor. *EMBO J.* 21: 135–144.
- Granek, J. A., and P. M. Magwene, 2010 Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet.* 6: e1000823.
- Granek, J. A., Ö. Kayıkçı, and P. M. Magwene, 2011 Pleiotropic signaling pathways orchestrate yeast development. *Curr. Opin. Microbiol.* 14: 676–681.
- Griffioen, G., P. Branduardi, A. Ballarini, P. Anghileri, J. Norbeck *et al.*, 2001 Nucleocytoplasmic distribution of budding yeast

- protein kinase A regulatory subunit Bcy1 requires Zds1 and is regulated by Yak1-dependent phosphorylation of its targeting domain. *Mol. Cell. Biol.* 21: 511–523.
- Guo, B., C. A. Styles, Q. Feng, and G. R. Fink, 2000 A *Saccharomyces* gene family involved in invasive growth, cell-cell adhesion, and mating. *Proc. Natl. Acad. Sci. USA* 97: 12158–12163.
- Hall, R. A., L. D. Sordi, D. M. Maccallum, H. Topal, R. Eaton *et al.*, 2010 CO(2) acts as a signalling molecule in populations of the fungal pathogen *Candida albicans*. *PLoS Pathog.* 6: e1001193.
- Hall-Stoodley, L., J. W. Costerton, and P. Stoodley, 2004 Bacterial biofilms: from the natural environment to infectious diseases. *Nat. Rev. Microbiol.* 2: 95–108.
- Halme, A., S. Bumgarner, C. Styles, and G. R. Fink, 2004 Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. *Cell* 116: 405–415.
- Honigberg, S. M., 2011 Cell signals, cell contacts, and the organization of yeast communities. *Eukaryot. Cell* 10: 466–473.
- Karunanithi, S., N. Vadaie, C. A. Chavel, B. Birkaya, J. Joshi *et al.*, 2010 Shedding of the mucin-like flocculin flo11p reveals a new aspect of fungal adhesion regulation. *Curr. Biol.* 20: 1389–1395.
- Kataoka, T., D. Broek, and M. Wigler, 1985 DNA sequence and characterization of the *S. cerevisiae* gene encoding adenylate cyclase. *Cell* 43: 493–505.
- Kuthan, M., F. Devaux, B. Janderová, I. Slaninová, C. Jacq *et al.*, 2003 Domestication of wild *Saccharomyces cerevisiae* is accompanied by changes in gene expression and colony morphology. *Mol. Microbiol.* 47: 745–754.
- Lambrechts, M. G., F. F. Bauer, J. Marmur, and I. S. Pretorius, 1996 Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohyphal differentiation in yeast. *Proc. Natl. Acad. Sci. USA* 93: 8419–8424.
- Lee, P., S.-M. Paik, C.-S. Shin, W.-K. Huh, and J.-S. Hahn, 2011 Regulation of yeast Yak1 kinase by PKA and autophosphorylation-dependent 14-3-3 binding. *Mol. Microbiol.* 79: 633–646.
- Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Liu, H., C. A. Styles, and G. R. Fink, 1996 *Saccharomyces cerevisiae* S288C has a mutation in FLO8, a gene required for filamentous growth. *Genetics* 144: 967–978.
- Lo, W. S., and A. Dranginis, 1998 The cell surface flocculin Flo11 is required for pseudohyphae formation and invasion by *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 9: 161–171.
- López, D., H. Vlamakis, and R. Kolter, 2010 Biofilms. *Cold Spring Harb. Perspect. Biol.* 2: a000398.
- Lorenz, M. C., and J. Heitman, 1997 Yeast pseudohyphal growth is regulated by GPA2, a G protein alpha homolog. *EMBO J.* 16: 7008–7018.
- Lorenz, M. C., X. Pan, T. Harashima, M. Cardenas, Y. Xue *et al.*, 2000 The G protein-coupled receptor gpr1 is a nutrient sensor that regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. *Genetics* 154: 609–622.
- Magwene, P. M., Ö. Kayıkçı, J. A. Granek, J. M. Reininga, Z. Scholl *et al.*, 2011a Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 108: 1987–1992.
- Magwene, P. M., J. H. Willis, and J. K. Kelly, 2011b The statistics of bulk segregant analysis using next generation sequencing. *PLOS Comput. Biol.* 7: e1002255.
- Malcher, M., S. Schladebeck, and H.-U. Mösch, 2011 The Yak1 protein kinase lies at the center of a regulatory cascade affecting adhesive growth and stress resistance in *Saccharomyces cerevisiae*. *Genetics* 187: 717–730.
- Markham, N. R., and M. Zuker, 2005 DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.* 33: W577–W581.
- McCusker, J. H., K. V. Clemons, D. A. Stevens, and R. W. Davis, 1994a Genetic characterization of pathogenic *Saccharomyces cerevisiae* isolates. *Genetics* 136: 1261–1269.
- McCusker, J. H., K. V. Clemons, D. A. Stevens, and R. W. Davis, 1994b *Saccharomyces cerevisiae* virulence phenotype as determined with CD-1 mice is associated with the ability to grow at 42 degrees C and form pseudohyphae. *Infect. Immun.* 62: 5447–5455.
- McDonough, K. A., and A. Rodriguez, 2012 The myriad roles of cyclic amp in microbial pathogens: from signal to sword. *Nat. Rev. Microbiol.* 10: 27–38.
- Michelmores, R. W., I. Paran, and R. V. Kesseli, 1991 Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. USA* 88: 9828–9832.
- Nishida, Y., F. Shima, H. Sen, Y. Tanaka, C. Yanagihara *et al.*, 1998 Coiled-coil interaction of N-terminal 36 residues of cyclase-associated protein with adenyl cyclase is sufficient for its function in *Saccharomyces cerevisiae* ras pathway. *J. Biol. Chem.* 273: 28019–28024.
- 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: 1564–1572.
- Palková, Z., 2004 Multicellular microorganisms: laboratory vs. nature. *EMBO Rep.* 5: 470–476.
- Palková, Z., and L. Váchová, 2006 Life within a community: benefit to yeast long-term survival. *FEMS Microbiol. Rev.* 30: 806–824.
- Pan, X., and J. Heitman, 1999 Cyclic amp-dependent protein kinase regulates pseudohyphal differentiation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 19: 4874–4887.
- Pan, X., and J. Heitman, 2002 Protein kinase A operates a molecular switch that governs yeast pseudohyphal differentiation. *Mol. Cell. Biol.* 22: 3981–3993.
- R Development Core Team, 2011 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- Reynolds, T. B., 2006 The Opi1p transcription factor affects expression of FLO11, mat formation, and invasive growth in *Saccharomyces cerevisiae*. *Eukaryot. Cell* 5: 1266–1275.
- Reynolds, T. B., and G. R. Fink, 2001 Bakers' yeast, a model for fungal biofilm formation. *Science* 291: 878–881.
- Schneeberger, K., S. Ossowski, C. Lanz, T. Juul, A. H. Petersen *et al.*, 2009 SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* 6: 550–551.
- Shao, Z., H. Zhao, and H. Zhao, 2009 DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 37: e16.
- Sherman, F., 2002 Getting started with yeast. *Methods Enzymol.* 350: 3–41.
- Stanford, D. R., M. L. Whitney, R. L. Hurto, D. M. Eisaman, W.-C. Shen *et al.*, 2004 Division of labor among the yeast Sol proteins implicated in tRNA nuclear export and carbohydrate metabolism. *Genetics* 168: 117–127.
- St'ováček, V., L. Váchová, M. Kuthan, and Z. Palková, 2010 General factors important for the formation of structured biofilm-like yeast colonies. *Fungal Genet. Biol.* 47: 1012–1022.
- Taylor, S. S., J. A. Buechler, and W. Yonemoto, 1990 cAMP-dependent protein kinase: framework for a diverse family of regulatory enzymes. *Annu. Rev. Biochem.* 59: 971–1005.
- Váchová, L., V. St'ováček, O. Hlaváček, O. Chernyavskiy, L. Stěpánek *et al.*, 2011 Flo11p, drug efflux pumps, and the extracellular

- matrix cooperate to form biofilm yeast colonies. *J. Cell Biol.* 194: 679–687.
- Verstrepen, K. J., T. B. Reynolds, and G. R. Fink, 2004 Origins of variation in the fungal cell surface. *Nat. Rev. Microbiol.* 2: 533–540.
- Wach, A., A. Brachat, R. Pöhlmann, and P. Philippsen, 1994 New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* 10: 1793–1808.
- Wu, J., T. Tolstykh, J. Lee, K. Boyd, J. B. Stock *et al.*, 2000 Carboxyl methylation of the phosphoprotein phosphatase 2A catalytic subunit promotes its functional association with regulatory subunits in vivo. *EMBO J.* 19: 5672–5681.
- Yi, S., N. Sahni, K. J. Daniels, K. L. Lu, T. Srikantha *et al.*, 2011 Alternative mating type configurations (a/α vs. a/a or α/α) of *Candida albicans* result in alternative biofilms regulated by different pathways. *PLoS Biol.* 9: e1001117.
- Yu, J., C. Wang, S. J. Palmieri, B. K. Haarer, and J. Field, 1999 A cytoskeletal localizing domain in the cyclase-associated protein, CAP/Srv2p, regulates access to a distant SH3-binding site. *J. Biol. Chem.* 274: 19985–19991.
- Zaman, S., S. I. Lippman, X. Zhao, and J. R. Broach, 2008 How *Saccharomyces* responds to nutrients. *Annu. Rev. Genet.* 42: 27–81.
- Zara, G., S. Zara, C. Pinna, S. Marceddu, and M. Budroni, 2009 Flo11 gene length and transcriptional level affect biofilm-forming ability of wild flor strains of *Saccharomyces cerevisiae*. *Microbiology* 155: 3838–3846.
- Zhang, Z., M. M. Smith, and J. S. Mymryk, 2001 Interaction of the E1A oncoprotein with Yak1p, a novel regulator of yeast pseudohyphal differentiation, and related mammalian kinases. *Mol. Biol. Cell* 12: 699–710.

Communicating editor: D. W. Threadgill

GENETICS

Supporting Information

<http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.142067/-/DC1>

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

Joshua A. Granek, Debra Murray, Ömür Kayrkçi, and Paul M. Magwene

Figure S1 There is extensive natural variation in *Saccharomyces cerevisiae* colony morphology. (A) Lab domesticates [BY4743] (Brachmann *et al.*, 1998) tend to form simple colonies, while diverse colony architectures more common among wild strains, such as those found from (B) Bertram Palm nectar [NCYC 3462] (Liti *et al.*, 2009) and (C) a distillery [YJM224] (Clemons *et al.*, 1994) , and clinical isolates from human (D) bile tube [YJM311] (McCusker *et al.*, 1994), (E) throat sputum [NCYC 3455] (Liti *et al.*, 2009) , and (F) vagina [NCYC 3471] (Liti *et al.*, 2009) . Scale bar is 1.1 mm.

Figure S2 The probability density estimates for distribution of adhesion to polystyrene ($\log(A_{590})$) by simple (black) and complex segregants (red). Complex colony segregants have significantly higher adhesion to polystyrene than simple colony segregants (Mann-Whitney $p=2.631 \times 10^{-12}$). The rug plot at bottom indicates values for individual segregants.

Figure S3 *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 with doxycycline incorporated at a range of concentrations. *CYR1-C* is much more effective at inducing complex morphology than *CYR1-S* Expression is shown increasing from left to right (doxycycline represses expression from the Tet vectors, so highest doxycycline corresponds with lowest expression). Numbers on lines correspond to the segregant number.

Figure S4 The probability density estimates for distribution of (A) *CYR1* and (B) *FLO11* expression values in simple (black) and complex segregants (red). Complex colony segregants have significantly higher expression of both *CYR1* and *FLO11* than simple colony segregants (Mann-Whitney p-values of 6.469×10^{-7} and 4.1×10^{-10} respectively). The rug plot at bottom indicates expression values for individual segregants.

Figure S5 Exogenous cAMP induces complex morphology in a segregant specific manner. Segregants s2,s4, s6, s9, and s12 were grown on YPD+water (-) and YPD+cAMP (+). The cAMP induces complex morphology in some but not all segregants.

Figure S6 Polystyrene adhesion by segregants correlates with *FLO11* expression levels. The correlation is most prominent for segregants with the *CYR1-S* allele. Almost all *CYR1-C* segregants are more adhesive than the *CYR1-S* segregants.

Figure S7 *Δflo8*, *Δsfl1*, and *Δmsn2* strains confirm a role for these transcription factor encoding genes in biofilm formation. *Δflo8* forms no biofilm while *Δmsn2* maintains minimal complex morphology. In the haploid background, *Δsfl1* appears to form stronger biofilms than the wildtype, which is supported by the gain of biofilm in the *Δsfl1* diploid a gain of biofilm relative to the wildtype diploid. MATa and MATα colonies of WT and knockouts showed the same phenotype, but only MATa colonies are shown. Colonies were imaged on day five of growth.

File S1
Primers Used In This Study

For each primer, gives the name, sequence, and how the primer was used. File S1 is a comma-separated-value (CSV) file available for download from the *Genetics* website at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.142067/-/DC1>.

File S2
QTL Details

For each statistically significant QTL, details the candidate gene, its distance from the QTL peak, coordinates of the QTL peak, and the rationale for selecting the candidate gene. The coordinates are based on *Saccharomyces* Genome Database (SGD) release r62, (release date: February 18, 2009). File S2 is a comma-separated-value (CSV) file available for download from the *Genetics* website at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.142067/-/DC1>. The columns are defined as follows.

Column	Label	Description
1	candidate	Candidate gene associated with the QTL peak
2	chromosome	Chromosome location of the QTL
3	peak_left	Chromosome coordinate of left end of peak (in bp)
4	peak_right	Chromosome coordinate of right end of peak (in bp)
5	peak_apex	Chromosome coordinate of peak apex (in bp)
6	candidate_distance	Distance of candidate gene from QTL peak apex (in bp)
7	note	Rationale for selecting candidate gene

File S3

***CYR1* Non-Synonymous Polymorphisms**

File S3 is a Portable Document Format (PDF) file available for download from the *Genetics* website at <http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.142067/-/DC1>. This file shows the non-synonymous polymorphisms in *CYR1-C* and *CYR1-S* relative to the S288c reference genome sequence. The functional domains of Cyr1p are indicated above the alignments. Amino acids that are ancestral (i.e. the s288c sequence is derived) are blue, novel amino acids (private to YJM311) are red, and those that are common with the pathogenic yeasts *Candida albicans* or *Ashbya gossypii* are yellow. Conserved positions are gray and gaps are indicated by a horizontal line.

Table S1 SNPs Genotyped by HRM

Gene	Chrom	Coordinate	Primer 1	Primer 2	SNP context
CYR1	10	425929	Cyr1snp2F	Cyr1snp2R	accaaagg[T/C]tggtgcgc
CYR1	10	428552	Cyr1snp12F	Cyr1snp12R	gaaaacct[T/C]gaactaac
CYR1	10	430641	Cyr1snp17F	Cyr1snp17R	tatcatga[G/A]cgagtagt
FLO11	9	394211	Flo11snp6F	Flo11snp6R	gcctgctc[A/C]agaataga
FLO11	9	393280	Flo11snp14F	Flo11snp14R	tttgaagt[T/C]tatgccac
FLO11	9	388882	Flo11snp30aF	Flo11downR2	cccactct[G/T]tctttcat

The SNPs that were assayed are identified by the combination of chromosome and chromosomal coordinate, and also by "SNP context", where the two alleles are uppercase, surrounded by square brackets, flanked by the eight bases immediately 5' and 3'. The SNP coordinates are based on *Saccharomyces Genome Database* (SGD) release r64, (release date: February 4, 2011).

Literature Cited

- Brachmann, C. B., A. Davies, G. J. Cost, E. Caputo, J. Li, *et al.*, 1998 Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**: 115–32.
- Clemons, K. V., J. H. McCusker, R. W. Davis, and D. A. Stevens, 1994 Comparative pathogenesis of clinical and nonclinical isolates of *Saccharomyces cerevisiae*. *J Infect Dis* **169**: 859–67.
- Liti, G., D. Carter, A. Moses, J. Warringer, L. Parts, *et al.*, 2009 Population genomics of domestic and wild yeasts. *Nature* .
- McCusker, J. H., K. V. Clemons, D. A. Stevens, and R. W. Davis, 1994 Genetic characterization of pathogenic *Saccharomyces cerevisiae* isolates. *Genetics* **136**: 1261–9.