Investigating genetic determinants of phenotypic variation in natural isolates of *Saccharomyces cerevisiae* 

by 

Hana Lee 

A dissertation submitted in partial satisfaction of the requirements for the degree of 

Doctor of Philosophy 

in 

Molecular and Cell Biology 

and the Designated Emphasis 

in 

Computational and Genomic Biology 

in the 

Graduate Division 

of the 

University of California, Berkeley 

Committee in charge: 

Professor Rachel B. Brem, Chair 

Professor Michael B. Eisen 

Professor Craig T. Miller 

Professor Yun S. Song 

Fall 2012
Investigating genetic determinants of phenotypic variation in natural isolates of *Saccharomyces cerevisiae*

© 2012

by Hana Lee
Abstract

Using population genomics to investigate genetic determinants of phenotypic variation in natural isolates of Saccharomyces cerevisiae

by

Hana Lee

Doctor of Philosophy in Molecular and Cell Biology
Designated Emphasis in Computational and Genomic Biology
University of California, Berkeley
Professor Rachel B. Brem, Chair

The causal link between genotype and phenotype is one of the fundamental principles of modern biology; yet there remain significant challenges to successfully identifying and validating the effect of a specific genetic variant on an organism. Much of the tremendous diversity observed in nature, even among individuals of the same species, remains unexplained. Here, we used the model eukaryote, Saccharomyces cerevisiae, to investigate naturally occurring variation and employ a candidate gene approach, through a combination of genome sequence analysis and mining functional gene annotations, to identify genetic determinants of the phenotypes observed. First, we took a look at morphological variation, a major source of biological diversity, in an environmental isolate of S. cerevisiae and found that its allele of CDC28 underlies multiple phenotypes: linearly arranged spores after meiosis, elongated cell shape during mitosis, and branching filaments during filamentous growth. Second, we studied a wild yeast population, using a comparative transcriptomics approach, which revealed divergence in iron metabolism that exhibited itself as slow growth in a high iron environment. We again identified two of the genetic determinants, YAP5 and CCC1, both essential for resistance to iron toxicity, that contribute to the phenotype and show evidence that genes involved in iron homeostasis have undergone non-neutral evolution. Our work illustrates the viability of using genomic data to successfully predict the genes responsible for phenotypes of interest as well as the power of yeast a model system for investigating natural variation.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ii

CHAPTER 1: Identifying the genetic determinants of phenotypic variation 1

CHAPTER 2: Natural variation in *CDC28* underlies morphological phenotypes in an environmental yeast isolate 4

  *Introduction* 4
  *Materials and Methods* 5
  *Results* 7
  *Discussion* 10
  *Figures and Tables* 12

CHAPTER 3: Divergence in iron metabolism in a wild yeast population 21

  *Introduction* 21
  *Materials and Methods* 22
  *Results* 24
  *Discussion* 27
  *Figures and Tables* 28

REFERENCES 39
ACKNOWLEDGMENTS

First, I would like to thank Dr. Rachel Brem for her considerable mentorship and support, as well as the members of my thesis committee, Dr. Michael Eisen, Dr. Craig Miller, Dr. Yun Song, and Dr. Adam Arkin, who provided numerous advice and suggestions over the course of my graduate studies.

The research presented in this dissertation also benefited from discussion and input from many people, both here at UC Berkeley and elsewhere. I owe a great deal to the fellow members of the Brem laboratory, including Dr. Charles Denby, Dr. Oh Kyu Yoon, Yulia Mostovoy, and Jeremy Roop. I have also benefited from discussions with Dr. Charles Ellison and Dr. Chris Ellison, who have regularly attended our group meetings. I also wish to particularly thank Dr. Jasper Rine and the members of his laboratory, for lending their expertise and sharing equipment, and Dr. Paul Magwene for his contributions to the publication that forms the second chapter of this dissertation. Dr. Daniel Lew and Dr. Liam Holt have also provided useful discussion during the preparation for publication.

I would also like to thank the Paul and Daisy Soros Foundation for sponsoring part of my graduate education through the Paul and Daisy Soros Fellowship for New Americans. The research presented in this dissertation was made possible through funding for my advisor, Dr. Rachel Brem, from the Ellison Medical Foundation and the National Institutes of Health.
CHAPTER 1

Identifying the genetic determinants of natural phenotypic variation

Fourty years ago, Francis Crick stated the central dogma of molecular biology that information encoded in DNA, through the intermediary of RNA, becomes translated into the amino acid sequence of proteins (Crick, 1970). Such sequence then determines protein structure and function, the molecular basis from which the biological traits of an organism emerge. While the scientific advances of the intervening decades have complicated our understanding of this simple model, the underlying hypothesis has remain unchanged: genotype creates phenotype. Much of the astounding biological diversity we observe—not only among different taxa separated by considerable evolutionary time, but even among individuals within the same species—arises from changes in DNA sequence.

In theory, simply knowing the genetic variants that distinguish one organism from another should give us sufficient information to explain inherited phenotypic differences. In practice, identifying what genes are responsible for a trait of interest, much less being able to predict how a particular sequence variant might affect an organism’s phenotype, is still very much a challenge. For example, the sequencing of the human genome ten years ago promised to revolutionize medicine, but the current reality is that the genome sequence alone is not yet sufficient for a complete understanding of the connection between genotype and phenotype. We have far to go before we can accurately describe the genetic basis of heritable human diseases, much less the full spectrum of biological diversity (Manolio et al., 2009; Lander, 2011).

There are several theoretical questions about the genetic architecture of complex traits and to what extent any generalizable principles can be found. For example, are most physiologically important phenotypes mediated primarily by a few genes of large effect or many genes of small effect? To what extent do the genetic determinants of a polygenic trait have epistatic interactions, and do these interactions tend to be additive or nonlinear? How pervasive is pleiotropy, in which a genetic variant affects multiple phenotype and how does it constrain evolutionary trajectories? Since cis-regulatory variation is less likely to be highly pleiotropic, are there differences in the frequency of cis-acting versus trans-acting variants that mediate regulatory changes between populations or species? While there have been many large-scale quantitative trait loci (QTL) analyses and genome-wide association studies (GWAS) in model organisms, from Drosophila to Arabidopsis, these questions still remain topics of debate in the field (Mackay, 2001; Wray, 2007; Flint and Mackay, 2009; Mackay et al., 2009).

Perhaps the most pressing challenge of all is to successfully identify a functional genetic variant that is causative for the phenotype of interest. Researchers have turned to computational approaches, particularly in the field of human genetic variation, that focus on predicting functional genetic variants a priori from the sequence information available. For example, there have been several attempts at developing prediction
methods for to detect which amino acid substitutions will have deleterious effects on protein function (Ng and Henikoff, 2001; 2006). Similarly, other researchers have invested their efforts in developing automated predictions of candidate genes, mining various sources of data, such as gene function annotations, definitions of biochemical pathways and regulatory networks, or searches of curated literature (Tiffin et al., 2005; Teber et al., 2009; Gefen et al., 2010). The primary drawback to all of these methods lie in their high false positive rates. However, the exponential rise in the amount of biological information, driven by the decreasing costs of sequencing, increases the power of these computational approaches.

In this respect, the budding yeast, *Saccharomyces cerevisiae*, has two main advantages as a model for studying natural variation and making genotype-phenotype associations. First of all, the wealth of genetic and biochemical research conducted in this laboratory model means that there are many large-scale experimental data sets available for further analysis as well as gene annotations for molecular function that is more comprehensive than in most other species. Secondly, the relative ease with which one can do site-directed mutagenesis, thanks to the high rate of homologous recombination, enables geneticists to go beyond prediction and association to directly test the effect of the implicated allele on phenotype. Studying the landscape of natural genetic variation in such an organism may provide insights that are applicable to other species.

While yeast has long been one of the most well-studied model organisms for molecular biology, the bulk of this research has utilized one of a few laboratory strains, which are phenotypic outliers and do not represent the full extent of diversity within the species. It has only been within the past twenty years that geneticists have begun to recognize the advantages of studying natural isolates of yeast. In addition to the multitude of domesticated strains for various types of fermentation and backing, *S. cerevisiae* can also exist as an opportunistic pathogen or in the wild. This diversity of lifestyles and environments make yeast an attractive model not only for addressing questions in quantitative genetics but for understanding the genetics of local adaptation (Landry et al., 2006a; Replansky et al., 2008).

Surveys of the genetic and phenotypic diversity in *S. cerevisiae* have become more common and larger in scale, with the development of increasingly high-throughput methods for sequencing and phenotyping. A number of population genetic analyses have described the population structure and landscape of polymorphisms in the species (Winzeler et al., 2003; Fay and Benavides, 2005; Schacherer et al., 2007), culminating in the release of whole-genome sequences for 36 isolates of *S. cerevisiae* (Liti et al., 2009). Gene expression profiling across panels of natural isolates have been used extensively to identify variation in the transcriptional response to different environmental stresses, such as copper sulfate (Fay et al., 2004), nitrogen limitation, and anaerobic conditions (Landry et al., 2006b). More recently, Warringer et al. carried out high-throughput phenotyping of the isolates sequenced in (Liti et al., 2009) under 200 different growth conditions to provide a catalog of the phenotypic variation in the
species (Warringer et al., 2011). However, identifying the specific genetic determinants that are causative for a phenotype of interest remains no easy task.

Two landmark studies used QTL mapping to identify the genetic determinants of complex trait variation between the laboratory type strain, S288c, and a natural isolate. Steinmetz et al. mapped thermotolerance observed in pathogenic strains of S. cerevisiae using the segregant progeny from a cross between S288c and a homozygous derivative of a clinical isolate (Steinmetz et al., 2002). Another QTL mapping, undertaken by Brem et al., treated gene expression as a quantitative trait and mapped variation in transcription using segregants from a cross between an S288c derivative and a vineyard isolate (Brem et al., 2002). This approach, called expression QTL (eQTL) analysis, has considerable power in elucidating gene regulatory networks in addition to differentiating between cis- and trans-acting variants. Other QTL and association studies have followed in their wake and achieved varying degrees of success in identifying the causative variants underlying phenotypes of interest (Gerke et al., 2006; Muller et al., 2011), but the genetic underpinnings of the tremendous diversity in yeast remain largely unknown.

The resequencing of 36 natural isolates has made possible some recent advances in our ability to interrogate the genetic determinants of natural phenotypic variation in S. cerevisiae. For example, Cubillos et al. took a QTL mapping approach on segregants derived from crosses between isolates from different S. cerevisiae populations for multiple environmental conditions chosen to mimic various ecological niches. The study revealed that most of these growth phenotypes were highly complex and the QTLs detected were frequently dependent on genetic background (Cubillos et al., 2011). In another approach, Jelier et al. used an algorithm to predict the effect of nonsynonymous and insertion-deletion polymorphisms on phenotype, using a combination of genome sequence alignments and experimental data from gene deletion screens. They successfully validated their method through high-throughput measurements of growth under different environmental conditions, thus highlighting the power of yeast as a model for understanding genetic variation (Jelier et al., 2011).

In the subsequent two chapters of this thesis, I combine genomic sequence analysis with the wealth of functional annotations available in yeast to predict candidate genes that underlie divergent phenotypes. In the second chapter, I investigate morphological variation observed in an environmental isolate and identify the main genetic determinant that underlies the phenotypes. In the third chapter, I report divergence in iron homeostasis between a wild yeast population and vineyard isolates and find two genes annotated for roles in response to iron toxicity contribute to the phenotype.
CHAPTER 2

Natural variation in \textit{CDC28} underlies morphological phenotypes in an environmental yeast isolate


INTRODUCTION

Natural genetic variation in morphology has given rise to Darwin’s “endless forms most beautiful” across the tree of life, and the search for the molecular basis of morphological variation remains a major focus of evolutionary genetics. To this end, unbiased genome-wide mapping is often used to find loci that underlie trait variation within species. However, the potential for high cost and limited statistical power of genome-wide analyses has led many researchers to test for causal DNA variants only in a subset of candidate genes, those predicted from prior knowledge about the trait of interest (Neale and Savolainen, 2004; Risch, 2000; Yoo et al., 2010). Classically, the candidate gene approach has met with varied degrees of success, owing to limited genotypic data and lack of detailed functional information for many genes and pathways (Hirschhorn and Daly, 2005; McCarthy et al., 2003; Tabor et al., 2002; Todd, 2006). Recently, increases in the availability of genome sequences and functional genomic resources have renewed interest in hypothesis-driven molecular genetics, in which functional variants can be predicted a priori from DNA sequence (Moreno et al., 2008; Wang and Marcotte, 2010).

The cell morphology and growth habit of budding yeast have been the target of genetic studies in laboratory strains (Jorgensen et al., 2002; Moseley and Nurse, 2009; Ohya et al., 2005; Pruyne et al., 2004) and environmental isolates (Fidalgo et al., 2006; Granek and Magwene, 2010; Nogami et al., 2007; Yvert et al., 2003), but the genetic underpinnings of many naturally occurring yeast morphologies remain unknown. In the search for the genetic basis of varying traits in yeast, a recently reported population genomic data set (Liti et al., 2009) established a powerful test bed for the candidate gene approach. In the present work, we undertook a proof of concept for hypothesis-driven genetic dissection of natural variation, using yeast sporulation as a model system. Upon observing an unusual arrangement of meiotic spores in an environmental yeast isolate, we set out to mine genome sequence data and test predictions about the underlying causal basis of this developmental phenotype.
MATERIALS AND METHODS

Reciprocal hemizygote strain construction
All strains used are listed in Table 2.1. The Ivory Coast (NCYC110) and oak (YP5606) wild isolates were obtained from the National Center for Yeast Collections (NCYC) in both diploid HO and haploid ho form. Ivory Coast x oak hybrid strains (YHL118 and YHL121) were generated from single-cell matings of the haploid ho strains from the two parental backgrounds to each other. These hybrid strains were subsequently used to make reciprocal hemizygotes by deleting one allele of CDC28 with a URA3 cassette. The CDC28 locus was sequenced in each transformant to verify that only one allele was present and to identify the allele. CDC28 hemizygotes were generated from homozygous diploid parent backgrounds via the same method.

Genomic analysis
Whole-genome alignments for a panel of genetically divergent strains, including the Ivory Coast and oak isolates, were retrieved from the Saccharomyces Genome Resequecing Project database (Liti et al., 2009). The aligned sequences for the open reading frames of annotated protein-coding genes were analyzed for the presence of nonsynonymous single nucleotide polymorphisms that were unique to the Ivory Coast strain.

Linkage analysis
A homothallic hybrid was constructed by mating single spores of HO wild-type Ivory Coast and oak strains to each other. This hybrid (YHL047) was sporulated and 24 tetrads were dissected. One homothallic segregant from each tetrad was sequenced at CDC28 and analyzed for sporulation and mitotic growth phenotypes. Sporulation morphology was quantified as the proportion of linear tetrads observed by microscopy, with >600 tetrads analyzed in each sample. Linkage to CDC28 was assessed by a Wilcoxon signed-rank test to compare segregant phenotypes associated with the two CDC28 alleles. The percent variance due to CDC28 genotype was calculated as the R2 of the linear regression of the segregant phenotypes to their CDC28 genotypes.

Sporulation
Cultures were grown for at least five days on solid minimal sporulation media (Amberg et al., 2005), prior to imaging at 40X magnification.

Mitotic growth
Cultures were grown to mid-log phase (OD600 ~ 0.7) in liquid rich media (Amberg et al., 2005), prior to imaging at 40X magnification.

Filamentous growth

5
Cultures were grown overnight on solid rich media. Cells from a single colony were suspended in water and diluted to an OD600 of 0.01 before plating on nitrogen-limited media with 1% butanol to induce filamentous growth (Gimeno et al., 1992). Colonies were imaged at 10X after four days of growth.
RESULTS

In wild-type yeast, meiosis is induced by nutrient limitation, and the four haploid spores resulting from meiotic divisions form a tetrahedral structure called a tetrad. We surveyed meiotic phenotypes in a panel of fully sequenced, genetically diverse yeast isolates which fall into well-defined phylogenetic populations (Liti et al., 2009). Upon starvation, all strains exhibited tetrads with a tetrahedral shape (Supplementary Figure 2.S1), except for NCYC110, an Ivory Coast isolate, in which we frequently observed an alternate form where the spores were arranged linearly in the ascus or spore sac (Figure 2.1A). This observation was particularly striking given the tetrahedral tetrad form in DBVPG6044 (Supplementary Figure 2.S1), an isolate from Guinea which is only 4.8 x 10^-5 percent divergent from the Ivory Coast strain across protein-coding regions. Linear tetrads have previously been observed in laboratory yeast, but only rarely (Lindsey et al., 2010; Piccirillo and Honigberg, 2010; Thomas and Botstein, 1987), and were used to deduce the first centromere linkage in yeast (Hawthorne, 1955).

To identify candidate alleles underlying the naturally occurring linear tetrad form of the Ivory Coast isolate, we searched population genomic sequence data (Liti et al., 2009) for nonsynonymous coding variants unique to this strain. Of the six genes harboring such variants (CDC28, PAU3, SEC27, SIW14, RAS2 and TRM44), we considered CDC28 to be the best candidate based on its known role in cell morphology. CDC28 is a cyclin-dependent kinase that regulates polarization during cell division (Lew and Reed, 1993). CDC28 knockouts are inviable and conditional loss-of-function mutations confer both meiotic and mitotic defects (Benjamin et al., 2003; Giaever et al., 2002; Kitazono et al., 2003; Kitazono and Kron, 2002), but various CDC28 point mutations give rise to enhanced polarized growth (Ahn et al., 2001; Edgington et al., 1999). Furthermore, a laboratory strain null for a CDC28 binding partner, CLB2, phenocopied the tetrad morphology of the Ivory Coast isolate (Figure 2.1B). The Ivory Coast strain harbored a unique serine-to-phenyalanine amino acid change at residue 79 in the CDC28 sequence, an amino acid forming part of a region known in the human homolog, Cdk1 (Cdc2), to be critical for Cdk-cyclin interactions (Jeffrey et al., 1995; Otyepka et al., 2006). On this basis, we hypothesized that the Ivory Coast allele of CDC28 was a genetic determinant of the linear tetrad phenotype.

To test this hypothesis, we first crossed the Ivory Coast strain to YPS606, a strain isolated from the exudate of North American oak trees, which has high sporulation efficiency (Gerke et al., 2006) and normal tetrad form (Figure 2.1). We collected haploid recombinant progeny from this cross and quantified the proportion of linear tetrads in each progeny strain. The results, shown in Figure 2.2, revealed strong linkage between the linear tetrad phenotype and genotype at CDC28 (Wilcoxon p = 3.5 x 10^-5). Consistent with the phenotypes of the parent strains, the Ivory Coast allele of CDC28 was associated with a high proportion of linear tetrads relative to the effect of the oak allele (Figure 2.2). Interestingly, segregants that bore the Ivory Coast allele at CDC28 did not all phenocopy the Ivory Coast parent: the frequency of linear tetrads varied across this strain set, reflecting the action of additional modifier loci segregating in the
cross (Figure 2.2). In contrast, across segregants bearing the oak allele at $CDC28$, proportions of linear tetrads were tightly distributed near zero (Figure 2.2). In an analysis of all segregants, 73% of variance in the linear tetrad phenotype was explained by the $CDC28$ locus. We conclude that the $CDC28$ locus is the major genetic determinant of this trait in the cross, with additional epistatic modifiers manifesting only in the presence of the Ivory Coast allele of $CDC28$ that together give rise to the phenotype of the Ivory Coast parent strain.

To confirm the identity of $CDC28$ as the major causative gene underlying the linear tetrad phenotype, we generated reciprocal hemizygotes (Steinmetz et al., 2002) for $CDC28$ in a diploid hybrid formed from a mating between the Ivory Coast and oak parents. This strategy produced two hybrid strains genetically identical to one another at all loci except $CDC28$, at which one hemizygote strain bore only the Ivory Coast allele and the other strain only the oak allele. Inducing meiosis in the wild-type Ivory Coast x oak hybrid yielded a low proportion of linear tetrads (Figures 2.1 and 2.3), indicating that the genetic determinants of the linear tetrad phenotype in the Ivory Coast strain act in a largely recessive manner. Consistent with predictions from linkage analysis, the hemizygote bearing only the Ivory Coast allele of $CDC28$ had a relatively high proportion of linear tetrads when sporulated, and linear tetrads were rare in sporulated cultures of the hemizygote bearing the oak allele of $CDC28$ (Figure 2.3). We conclude that $CDC28$ is a major causative gene for the high frequency of linear tetrads in the Ivory Coast strain.

The evidence for dominance by the oak allele of $CDC28$ in the genetics of the linear tetrad trait (Figure 2.3) suggested that the Ivory Coast allele at this locus was likely hypomorphic. However, given that the oak x Ivory Coast hybrid strain produced a modest proportion of linear tetrads when sporulated (Figures 2.1 and 2.3), we further hypothesized that there was a subtle effect of haploinsufficiency at $CDC28$ in the hybrid relative to the oak parent. Under this model, we expected that a reduction in $CDC28$ dose would also give rise to an increased frequency of linear tetrads in either parental strain background. To test this notion, we deleted one copy of $CDC28$ in the oak homozygous parent and, separately, in the Ivory Coast parent. In each case, the hemizygote produced more linear tetrads than did its respective homozygote (Figure 2.3), with a more dramatic effect in the Ivory Coast background. Thus, the functional dose of $CDC28$ correlated inversely with the proportion of linear tetrads; the Ivory Coast allele of $CDC28$ acted as a partial loss of function, and haploinsufficiency in $CDC28$ in either strain increased the frequency of linear tetrads. Notably, the hemizygote bearing a given allele in the parental background did not phenocopy the respective hemizygote in the hybrid (Figure 2.3), consistent with the evidence from linkage analysis (Figure 2.2) for additional variant loci as modifiers to $CDC28$.

In laboratory yeast, experimentally inducing elongated morphology during mitotic growth gives rise to a linear tetrad phenotype upon sporulation (Lindsey et al., 2010; Thomas and Botstein, 1987), and a laboratory strain null for the CDC28 binding partner CLB2 grows as elongated cells in rich medium (Ahn et al., 1999). We hypothesized that the
natural variant of \textit{CDC28} in the Ivory Coast strain would also confer elongated cell shape in mitotically budding cells. Consistent with this prediction, the Ivory Coast parent and hemizygotes bearing the Ivory Coast allele of \textit{CDC28} showed elongated cell morphology during mitotic growth, in contrast to the round cells of the oak parent and the hemizygotes with the oak allele of \textit{CDC28} (Figure 2.4). Furthermore, in the progeny of the cross between the oak and Ivory Coast strains, segregants with hyperpolarized cells in rich media also produced linear tetrads when sporulated (Supplementary Figure 2.3). Interestingly, in contrast to our analysis of tetrads (Figure 2.3), we found no evidence for haploinsufficiency in \textit{CDC28} dose as a determinant of cell shape: oak homozygotes with a single copy of \textit{CDC28} still grew as round cells, as did the wild-type hybrid between the oak and Ivory Coast strains (Figure 2.4). We conclude that the Ivory Coast allele of \textit{CDC28} is sufficient to confer elongated cell morphology, reflecting a relationship between hyperpolarized mitotic growth and linear tetrads during sporulation.

Many strains of yeast, when starved for nitrogen, undergo unipolar budding, which generates filaments of linked cells (Gimeno et al., 1992). Given the pronounced filamentous phenotype under nitrogen limitation in a laboratory strain mutant for CLB2, the binding partner of \textit{CDC28} (Ahn et al., 1999), we hypothesized that the Ivory Coast allele of \textit{CDC28} would drive a similar phenotype. In accord with this expectation, the Ivory Coast isolate, when starved for nitrogen, grew as long, branching filaments composed of elongated cells growing from the colony edge, whereas the oak strain showed a more modest response with non-branching filaments composed of round cells (Figure 2.5). Reciprocal hemizygote analysis confirmed that the Ivory Coast allele of \textit{CDC28} was sufficient to confer highly branched, elongated filaments under nitrogen limitation (Figure 2.5). Again, we observed no evidence for haploinsufficiency of \textit{CDC28} dose as a determinant of filamentous morphology: hemizygous strains generated from the homozygous parent backgrounds resembled their respective parents, and the wild-type hybrid between the oak and Ivory Coast strains resembled the oak parent (Figure 2.5). Thus, the Ivory Coast allele of \textit{CDC28} underlies the dramatic filamentation behavior in this environmental isolate, mirroring its effects on tetrad and single-cell morphology.
DISCUSSION

Many variations in yeast cell form and growth habit have been reported in environmental isolates (Casalone et al., 2005; Cavalieri et al., 2000; Dengis et al., 1995; Granek and Magwene, 2010; Nogami et al., 2007; St'oviček et al., 2010), with the molecular basis known for only a small subset of these traits. We have shown that natural genetic variation at a single essential gene, CDC28, can lead to striking changes in yeast cell shape during meiotic and mitotic division. The rod-like tetrads, elongated cell shape, and hyperfilamentous growth of the environmental Ivory Coast isolate studied here can be attributed to variation in CDC28. Previous reports have implicated loss of function in Cdc28/Clb2 activity in hyperpolarized cell division and filamentous growth (Ahn et al., 1999; 2001; Edgington et al., 1999; Lew and Reed, 1993), but temperature-sensitive mutations available in laboratory strains have not been well-suited to the study of CDC28 loss of function under starvation conditions. By contrast, our ability to map starvation-induced sporulation and filamentation phenotypes to CDC28 highlights the utility of analysis of natural genetic variation when drawing relationships between genes and phenotypes.

The molecular mechanism by which the Ivory Coast allele of CDC28 impacts tetrad and cell morphology remains an open question. The phenotypic similarity between the Ivory Coast strain and the Δclb2 lab strain, and the S79F change in the former in a known cyclin-interaction domain of Cdc28, suggest that the Ivory Coast allele attenuates the binding of CDC28 to CLB2. We favor a model in which elongated cell shape is the proximate cause of the linear tetrad morphology and dramatic filamentation phenotype attributable to the Ivory Coast allele, in light of the relationship between these traits when cell shape is manipulated in the laboratory (Lindsey et al., 2010; Thomas and Botstein, 1987). With respect to sporulation, one compelling idea for the underlying mechanism is that, although meiotic cell division is ordered in yeast (Hawthorne, 1955), spores arrange passively in the ascus as a function of cell shape before meiosis (Piccirillo and Honigberg, 2010). Among the most notable aspects of the Ivory Coast allele of CDC28 are its incomplete penetrance and its sensitivity to genetic modifiers, for which the Ivory Coast strain’s unique coding alleles of the morphology regulator RAS2 and other genes represent strong candidates. As such, our data indicate that the effect of the Ivory Coast variant of CDC28 on linear tetrad formation is highly sensitized both to stochastic variation in cell shape and to genetic differences between yeast strains.

The evolutionary pressures driving variation in yeast morphology and growth habit in the wild are as yet almost entirely unknown, although several morphological characters are well-characterized targets of artificial selection in beer and wine yeasts (Verstrepen et al., 2003). Given the origin of the Ivory Coast strain as an isolate from ginger beer (Liti et al., 2009), its elongated cell and tetrad forms may be associated with a fitness benefit in the wild or during domestication; alternatively, these phenotypes, and the variant in the essential cell cycle regulator CDC28, may represent a rare genetic defect destined to be eliminated by purifying selection. In either case, our discovery of CDC28 as the major determinant underscores the power of combining dense population genomic...
sampling with a candidate gene/candidate network approach. With the continued growth of population genomic sequence compendia, and increasing knowledge of gene function and genetic interactions, this strategy holds promise for application to many organisms and traits.
<table>
<thead>
<tr>
<th>STRAIN NAME</th>
<th>PARENTS</th>
<th>GENOTYPE</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCYC110</td>
<td>MATa/α</td>
<td></td>
<td>NCYC</td>
</tr>
<tr>
<td>YPS606</td>
<td>MATa/α</td>
<td></td>
<td>NCYC</td>
</tr>
<tr>
<td>NCYC110 MATa</td>
<td></td>
<td>MATa ura3Δ::kanMX hoΔ::hphMX</td>
<td>NCYC</td>
</tr>
<tr>
<td>NCYC110 MATα</td>
<td></td>
<td>MATa ura3Δ::kanMX hoΔ::hphMX</td>
<td>NCYC</td>
</tr>
<tr>
<td>NCYC110 MATa/α</td>
<td></td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX</td>
<td>NCYC</td>
</tr>
<tr>
<td>YPS606 MATa</td>
<td></td>
<td>MATa ura3Δ::kanMX hoΔ::hphMX</td>
<td>NCYC</td>
</tr>
<tr>
<td>YPS606 MATα</td>
<td></td>
<td>MATa ura3Δ::kanMX hoΔ::hphMX</td>
<td>NCYC</td>
</tr>
<tr>
<td>YPS606 MATa/α</td>
<td></td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX</td>
<td>NCYC</td>
</tr>
<tr>
<td>G30045</td>
<td>Σ1278b</td>
<td>MATa ura3Δ0/ura3Δ0 his3Δ0/his3Δ0 clb2Δ::kanMX/ clb2Δ::kanMX</td>
<td>(Drees:2005hc)</td>
</tr>
<tr>
<td>YHL047</td>
<td>NCYC110 x YPS606</td>
<td>MATa/α HO/HO</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL118</td>
<td>NCYC110 MATa x YPS606 MATα</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL121</td>
<td>NCYC110 MATα x YPS606 MATa</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL169</td>
<td>NCYC110 MATa/α</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-NCYC110</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL170</td>
<td>YPS606 MATa/α</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-YPS606</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL171</td>
<td>YPS606 MATa/α</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-YPS606</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL172</td>
<td>NCYC110 MATa/α</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-NCYC110</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL173</td>
<td>YHL118</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-YPS606</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL174</td>
<td>YHL118</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-YPS606</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL175</td>
<td>YHL118</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-NCYC110</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL176</td>
<td>YHL121</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-NCYC110</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL177</td>
<td>YHL121</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-NCYC110</td>
<td>this paper</td>
</tr>
<tr>
<td>YHL178</td>
<td>YHL121</td>
<td>MATa/α ura3Δ::kanMX/ura3Δ::kanMX hoΔ::hphMX/ hoΔ::hphMX cdc28Δ::URA3/CDC28-NCYC110</td>
<td>this paper</td>
</tr>
</tbody>
</table>
**Supplementary Figure 2.S1 — Spore morphologies of non-mosaic strains.** Each panel shows a representative micrograph taken at 40X magnification after five days of growth on solid minimal sporulation media. The Ivory Coast and oak strains used for subsequent analysis are NCYC110 and YPS606, respectively. The Guinea isolate most closely related to the Ivory Coast strain is DBVPG6044. Sake strain K11 is not shown due to extremely low sporulation efficiency.
Figure 2.1 — Linear tetrad morphology observed in an Ivory Coast isolate of *S. cerevisiae*. Each panel shows a micrograph of yeast culture taken after five days of growth on solid minimal sporulation media. Examples of linear tetrads are indicated with arrowheads. (A) Ivory Coast and North American oak strains. (B) Σ1278b laboratory strain bearing a deletion in CLB2. (C) Ivory Coast x oak hybrid.
Figure 2.2 — The linear tetrad phenotype links to the CDC28 locus. Each column reports a distribution of proportions of linear tetrads across cultures of homozygote, homothallic yeast strains grown for five days of growth on solid minimal sporulation media. For each distribution, the median is reported as a thick horizontal line, the 25% quantile is shown as a box, and the extremes are shown as thin horizontal bars.

Parent: homozygotes derived from wild isolates (Liti et al., 2009); distributions represent sporulation phenotypes across three biological replicates. Segregants, Ivory Coast allele: F1 progeny from a cross between Ivory Coast and oak strains bearing the Ivory Coast allele of CDC28 ($n = 12$). Segregants, oak allele: F1 progeny bearing the oak allele of CDC28 ($n = 12$).
Figure 2.3 — Natural variation in CDC28 is causative for the linear tetrad phenotype. (A) Proportion of linear tetrads in hemizygote and wild-type strains. Each
column reports data from three biological replicates (parental wild-type strains) or independently generated isogenic strains (hybrids and hemizygotes). Bar heights report means, and error bars represent one standard deviation. (B) Micrographs of CDC28 reciprocal hemizygote strains in the oak x Ivory Coast hybrid background. (C) Micrographs of CDC28 hemizygote strains in homozygous oak and Ivory Coast parental backgrounds. Examples of linear tetrads are indicated with arrowheads. All micrographs were taken after five days of growth on solid minimal sporulation media.
Figure 2.4 — CDC28 genotype affects cell shape during mitotic growth. Each panel shows a representative micrograph of culture grown to mid-log phase in rich media. (A) Ivory Coast and North American oak strains. (B) Oak x Ivory Coast wild-type hybrid strain. (C) CDC28 reciprocal hemizygotes in the oak x Ivory Coast hybrid background. (D) CDC28 hemizygotes in homozygous oak and Ivory Coast parental backgrounds.
Supplementary Figure 2.S2 — Elongated cell morphology during mitotic growth segregates with linear tetrad morphology during sporulation. Segregants from the homothallic Ivory Coast x oak hybrid were sporulated and, separately, were grown in rich media, and representative micrographs were taken at 40X magnification.
Figure 2.5 — CDC28 genotype affects morphology of the filamentous growth response. Each panel represents a representative micrograph of colonies after four days of growth on solid nitrogen limitation media. (A) Ivory Coast and North American oak strains. (B) Oak x Ivory Coast wild-type hybrid. (C) CDC28 reciprocal hemizygotes in the oak x Ivory Coast hybrid background. (D) CDC28 hemizygotes in homozygous oak and Ivory Coast parental backgrounds.
CHAPTER 3

Divergence in iron metabolism in a wild yeast population

INTRODUCTION

A main goal of comparative genomics is to understand how and when regulatory changes between populations give rise to macroscopic trait differences. Gene expression variation is widespread, with species comparisons routinely detecting regulatory change at more than half of the genes in a given genome (Hodgins-Davis and Townsend, 2009; Wilson and Odom, 2009; Dowell, 2010). Landmark studies have traced trait differences between populations to divergence in gene regulation (Shapiro et al., 2004; Hanikenne et al., 2008; Rebeiz et al., 2009; Tung et al., 2009; Wittkopp et al., 2009; Booth et al., 2010; Loehlin and Werren, 2012), but for the vast majority of cases, the relationship between gene expression changes and phenotype remains unknown; methods to interpret the phenotypic relevance of expression variation are at a premium in the modern literature.

In this work, we set out to use wild yeast as a model system for the study of regulatory and phenotypic variation between populations. Nectar of the flowers of Malaysian bertam palm trees hosts a yeast community whose fermentation is consumed by penta-tailed tree shrews, a primary pollinator of the trees (Wiens et al., 2008). \textit{S. cerevisiae} strains collected from this nectar are reproductively isolated from the rest of the species (Cubillos et al., 2011), and we sought to investigate the evolution of this unique yeast population, using transcriptional profiling to generate hypotheses about trait divergence. The results led us to the discovery of divergent iron homeostasis and iron toxicity resistance behaviors in the Malaysian strains, and to a dissection of the genetic and evolutionary basis of these macroscopic phenotypes.
MATERIALS AND METHODS

**RNAseq library preparation**

All strains used are listed in Supplementary Table 3.S1. The three Malaysian (UWOPS03.461.4, UWOPS05.217.3, UWOPS05.227.2) and two Wine/European (BC187, RM11-1) isolates used in this study were obtained from the National Collection of Yeast Cultures (NCYC). For analysis of cis-regulatory variation in Supplementary Figure 3.S1, four Malaysian x Wine/European hybrids were generated through single-cell matings between Malaysian and Wine/European strains: two UWOPS03.461.4 x RM11-1 isogenic hybrids, one UWOPS03.461.4 x BC187 hybrid and one UWOPS05.217.3 x BC187 hybrid.

Parental and hybrid strains were grown to mid-log phase (0.65-0.72 OD$_{600}$) in YPD medium (Amberg et al., 2005). Total RNA was extracted from cells using the hot phenol method (Ausubel, 2002), and genomic DNA was removed with Turbo DNase I (Life Technologies). Libraries were prepared as in (Mortazavi et al., 2008), and sequenced on Illumina 2G Genome Analyzer and HiSeq 2000 machines with 100 bp paired-end reads.

**RNA-seq**

RNA-seq reads from Malaysian and wine/European isolates were mapped against the S288c reference genome ([http://www.yeastgenome.org](http://www.yeastgenome.org)) using Bowtie (Langmead et al., 2009). Single-nucleotide polymorphisms were called using SAMtools (Li et al., 2009) and used to amend the genome sequences downloaded from (Liti et al., 2009). These amended sequences were then used as references for re-mapping of RNA-seq reads, with only those mapping uniquely and containing no mismatches used for further analysis.

Read counts per gene were summed using custom Python scripts and normalized with the EDASeq package in R (Risso et al., 2011), using the upper quartile method for differences in library size between lanes and the loess method for differences in GC content within lanes. Differential gene expression ratios and p-values were calculated using DESeq, with correction for multiple testing using the Benjamini-Hochberg method implemented in the DESeq package (Anders and Huber, 2010).

**Expression Analysis**

To screen regulons for directional expression divergence between Malaysian and wine/European yeast, coregulated gene clusters were obtained from (Gasch et al., 2004) and GO terms were obtained from the Saccharomyces Genome Database ([http://www.yeastgenome.org](http://www.yeastgenome.org); April 2010 release). For each regulon, we tabulated all genes whose differential expression between the populations was significant at adjusted p-value < 0.05, and summed the sign of their log$_2$ fold changes. We then repeated this sign sum procedure on each of 10,000 random gene groups of the same size as the regulon of interest, drawn from the total set of genes with significant differential
expression. The proportion of such null gene groups with a sign sum as extreme as, or more extreme than, the sign sum of the real regulon was taken to be the nominal empirical p-value for directional expression divergence. P-values were corrected for multiple testing using the Benjamini-Hochberg method as implemented in the p.adjust() function in R.

To analyze iron-toxicity response genes in Figure 3.1 and Table 3.2, we used the set of genes reported to be upregulated in a Δccc1 laboratory strain, compared to wild-type, in 3 mM FeSO₄ from (Lin et al., 2011). Expression analysis was as described above and sequence analysis as below.

Growth assays
For each strain, a pre-culture was prepared by growth for four hours in YPD medium and diluted to 0.1 OD₆₀₀ by resuspension in distilled water. Experimental cultures were then established by diluting this pre-culture 1:10 in complete synthetic medium (CSM) (Amberg et al., 2005) or CSM supplemented with 5 mM FeSO₄ for a starting concentration of 0.01 OD₆₀₀ in a 96-well plate. Each strain in each condition was grown in triplicate. Growth was monitored via OD₆₀₀ measurements every 30 minutes in a Tecan Infinite F500 microplate reader. Doubling times were calculated from growth curves using the method of (Warringer et al., 2011) implemented in Python.

Reciprocal hemizygote strain construction
Malaysian x Wine/European hybrid strains (YHL243 and YHL247; see Supplementary Table 3.S1) were used to make reciprocal hemizygotes by deleting one allele of the gene of interest (either CCC1 or YAP5) with a URA3 cassette (Sikorski and Hieter, 1989). The targeted locus was sequenced in each transformant to verify that only one allele was present and to identify the allele. Two to four transformants were analyzed for each reciprocal hemizygote constructed.

Kₐ/Kₛ analysis
Multiple sequence alignments were generated for each gene using MUSCLE (Edgar 2004) from genome sequences of all non-mosaic strains as defined in (Liti et al., 2009), using uncorrected published sequences. The ratio of nonsynonymous to synonymous substitution rates, using the formula for Kₐ/Kₛ in (Nei and Gojobori, 1986), was calculated for each gene using SNPs fixed in and private to each non-mosaic yeast population as defined in (Liti et al., 2009). For each population, we calculated the mean of these Kₐ/Kₛ across all genes of the iron-toxicity response regulon as defined above. We then repeated this sign sum procedure on each of 10,000 random gene groups of the same size as the iron-toxicity response regulon, drawn from all genes for which SNP data were available. The proportion of such null gene groups with a sum as extreme as, or more extreme than, the sum of the real regulon was taken to be the empirical p-value for protein evolutionary rate.
RESULTS

To survey gene expression programs in wild yeast, we took a comparative approach using three homozygote Malaysian isolates and two homozygotes from a distinct, well-defined wine/European yeast population (Liti et al., 2009). We cultured each strain in standard rich medium and subjected each to transcriptional profiling by RNA-seq (Supplementary Table 3.5). We used the RNA-seq reads to verify and correct coding sequences from low-coverage genomes available for these strains (Liti et al., 2009), identifying a total of 46,367 single-nucleotide polymorphisms in coding regions divergent between the Malaysian and wine/European strains and fixed within each set of strains. Mapping to the amended genomes and analysis of the complete set of profiles revealed 601 genes differentially expressed between the populations.

We sought to focus on patterns of regulatory variation, which could be a signpost for macroscopic trait differences between Malaysian and wine/European yeast. For this purpose, we screened sets of functionally related genes for coherent expression change between the populations. Testing gene groups defined by Gene Ontology (http://www.yeastgenome.org) and those defined on the basis of co-regulation in classic analyses of laboratory yeast (Gasch et al., 2004) we identified eight GO terms and six regulons whose genes were predominantly upregulated, or predominantly downregulated, in Malaysian strains relative to wine/European strains (false discovery rate = 5%; Table 3.1). Among these were two sets of genes annotated in iron metabolism, for which the Malaysian population was associated with low expression: targets of AFT1, a transcription factor involved in iron homeostasis, and the iron ion transport Gene Ontology term GO:0006826. This largely non-overlapping pair of regulons comprises genes involved in iron uptake under conditions of iron starvation (Figure 3.1D).

Given the divergent expression of iron starvation genes between Malaysian and wine/European yeasts, we hypothesized that regulation of the response to excess iron might also be distinct between the populations. We examined a suite of genes upregulated by laboratory yeast in iron-toxic conditions, which includes components of the DNA damage and oxidative stress response pathways (Lin et al., 2011). Testing this regulon for expression differences between yeast populations again revealed coherent regulatory change, with the Malaysian strains in this case associated with high expression (resampling p = 0.011; Figure 3.1C). Thus, Malaysian yeast cultured in standard laboratory conditions repressed iron-uptake genes, and activated genes involved in resistance to iron toxicity, relative to wine/European isolates (Figure 3.1D).

We reasoned that the regulatory differences in iron metabolism between yeast populations could have phenotypic correlates in the context of growth during iron challenge. To test this, we measured growth rates of Malaysian and wine/European isolates, and hybrids formed by the mating of individuals from the two populations to one another, in high-iron conditions and in a standard-medium control. The results bore out our prediction, with Malaysian strains conferring slower growth than wine/European
isolates in high-iron medium (Figure 3.2). This defect acted as a recessive trait, as hybrid strains showed resistance to high-iron treatment on par with that of their wine/European parents (Figure 3.2). We detected no differences across strains in growth under iron starvation conditions (data not shown). We conclude that Malaysian yeast strains are uniquely sensitive to iron toxicity, corroborating our observations of regulatory divergence in iron-response pathways and reflecting the likely origin of the iron-sensitivity phenotype as a derived trait in the Malaysian population.

We next sought to dissect the genetic basis of iron sensitivity in Malaysian strains using a candidate gene approach. We expected that causative variants underlying the trait were likely to act in *trans* on the iron metabolism machinery, since measurements of allele-specific expression in Malaysian x wine/European hybrids showed no strong signals of *cis*-regulatory variation in iron metabolism genes (Supplementary Figure 3.S1). To identify candidate loci underlying variation in iron response, we searched the genome for non-synonymous coding polymorphisms unique to the Malaysian population in components of the iron metabolism gene network. Among the resulting gene hits, we focused on two known to be essential in laboratory yeast for growth in high-iron conditions (Li et al., 2001; 2008): the transcription factor *YAP5*, which activates genes in response to high iron concentrations, and *CCC1*, a transporter that sequesters iron ions in the vacuole. In *YAP5* we observed a glutamine-to-histidine substitution in the leuzine zipper DNA-binding domain and an arginine-to-glycine substitution near the N-terminus of the protein, both private to Malaysian strains; in *CCC1* we observed a glycine-to-arginine substitution near the N-terminus, again unique to Malaysian yeasts (Supplementary Figure 3.S2).

To evaluate the role of genetic variation at *YAP5* and *CCC1* in iron sensitivity, we used reciprocal hemizygote analysis (Steinmetz et al., 2002) for each gene as follows. In the diploid hybrid formed by mating a haploid Malaysian strain to a haploid wine/European, we knocked out each allele in turn of the gene of interest. The resulting pair of hybrid strains were isogenic to one another throughout the genome except at the hemizygous locus, where each strain harbored only one of the two alleles from the parent strains; any phenotypic differences between the strains of a hemizygote pair could thus be ascribed to variation at the manipulated site. We measured growth rates of each hemizygote in high-iron medium and a standard-medium control, and we compared growth between the strains of a given pair. The results, shown in Figure 3.2, revealed a 1.18-fold difference between strains bearing the Malaysian and wine/European alleles of *CCC1*, and a more modest 1.05-fold, though still significant, effect of variation in *YAP5*. In each case, as predicted, the Malaysian allele of the respective gene was associated with compromised growth in high-iron conditions (Figure 3.2). Additionally, strains hemizygous for *CCC1* grew more slowly than wild-type hybrids in high-iron medium regardless of which allele they harbored (Figure 3.2), reflecting haploinsufficiency at this gene with respect to iron resistance. We conclude that iron resistance is a genetically complex trait mediated in part by variants in *YAP5* and *CCC1*, in which Malaysian alleles act as losses of function, compromising growth in high-iron conditions.
We next aimed to investigate the evolutionary forces driving divergence in iron metabolism by Malaysian strains. Using single-nucleotide polymorphisms private to and fixed in each of five yeast populations characterized by (Liti et al., 2009), we estimated ratios of non-synonymous to synonymous substitution rates in coding regions and compared these rates between the regulon expressed during high-iron exposure (Lin et al., 2011) and the rest of the genome. The results, shown in Table 3.2, revealed a 1.6-fold elevation of protein evolutionary rate in iron metabolism genes of Malaysian strains, with no such effect detectable in any other yeast population. These results suggest that the sequences of iron metabolism genes exhibit a signature of non-neutral evolution in Malaysian yeast, lending further support to the notion of derived iron-response behaviors in the Malaysian population.
DISCUSSION

Against a backdrop of hundreds of comparative transcriptomic studies in many species, a continuing challenge in the field is to understand when and how regulatory divergence between populations is relevant for macroscopic traits. Addressing the question requires the inference of relationships between expression and phenotype, for which we have established iron metabolism in Malaysian yeast as a model system.

We have shown that Malaysian strains bear loss-of-function alleles in the iron-resistance genes *YAP5* and *CCC1*, and an excess of coding changes in many other iron metabolism genes. In addition, the Malaysian population perceives standard culture medium as a high-iron condition, activating iron-resistance expression programs which echo the upregulation of stress responses seen in other fungi under laboratory conditions (Tirosh et al., 2011; Martin et al., 2012). Taken together, these patterns support a model of specialization by Malaysian strains to a low-iron environment. Under this scenario, Malaysian yeasts would have accumulated genetic change in the iron homeostasis machinery as a consequence of relaxed selection on resistance to high iron, or an evolutionary trade-off to optimize fitness in conditions other than those studied here. In either case, our findings provide a case in which the recessive loss-of-function mutations that appear to litter wild yeast genomes (Zörgő et al., 2012) follow a compelling evolutionary logic. A broader involvement of additional metals is suggested by the growth defect of a Malaysian isolate in high-copper medium (Hodgins-Davis et al., 2012). The emerging picture is one in which the Malaysian yeast population has experienced unique evolutionary pressures on metal metabolism, highlighting the palm-flower niche of these yeasts as a driver of evolutionary change.

Our work in the wild yeast model makes clear that genomic signatures of regulatory change can be used as a powerful hypothesis generator for the study of trait variation. With the increased availability of functional-genomic resources in many taxa, this approach may prove broadly applicable in the study of species and population differentiation.
### FIGURES AND TABLES

**Supplementary Table 3.S1 — Strains used in this work.**

<table>
<thead>
<tr>
<th>STRAIN NAME</th>
<th>PARENTS</th>
<th>GENOTYPE</th>
<th>SOURCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>UWOPS03.461.4</td>
<td>MATα/α HO</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>UWOPS05.217.3</td>
<td>MATα/α HO</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>UWOPS05.227.2</td>
<td>MATα/α HO</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>BC187</td>
<td>MATα/α HO</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>UWOPS03.461.4 a/α</td>
<td>MATα/α ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>UWOPS05.217.3 a/α</td>
<td>MATα/α ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>UWOPS05.227.2 a/α</td>
<td>MATα/α ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>BC187 a/α</td>
<td>MATα/α ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>RM11-1</td>
<td>RM11-1a x RM11-1b MATα/α ura3 hoΔ:kanMX</td>
<td>(Brem:2002eu)</td>
<td></td>
</tr>
<tr>
<td>UWOPS03.461.4 a</td>
<td>MATα ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>UWOPS03.461.4 α</td>
<td>MATα ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>UWOPS05.217.3 a</td>
<td>MATα ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>UWOPS05.217.3 α</td>
<td>MATα ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>BC187 a</td>
<td>MATα ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>BC187 α</td>
<td>MATα ura3Δ:kanMX hoΔ:hpMX</td>
<td>NCYC</td>
<td></td>
</tr>
<tr>
<td>RM11-1a</td>
<td>MATα leu2 ura3 hoΔ:kanMX</td>
<td>(Brem:2002eu)</td>
<td></td>
</tr>
<tr>
<td>RM11-1b</td>
<td>MATα lys2 ura3 hoΔ:kanMX</td>
<td>(Brem:2002eu)</td>
<td></td>
</tr>
<tr>
<td>YHL057</td>
<td>UWOPS03.461.4 x RM11-1b MATα/α HO</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>YHL058</td>
<td>UWOPS03.461.4 x BC187 a MATα/α HO</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>YHL063</td>
<td>UWOPS05.217.3 x BC187 MATα/α HO</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>YHL065</td>
<td>UWOPS05.217.3 x BC187 MATα/α HO</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>YHL243</td>
<td>UWOPS03.461.4 a x BC187 a MATα/α ura3Δ:kanMX hoΔ:hpMX</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>YHL247</td>
<td>UWOPS03.461.4 a x BC187 α MATα/α ura3Δ:kanMX hoΔ:hpMX</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>YHL408</td>
<td>YHL243 MATα/α ura3Δ:kanMX hoΔ:hpMX yap5Δ::URA3/YAP5-WineEuropean</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>YHL409</td>
<td>YHL243 MATα/α ura3Δ:kanMX hoΔ:hpMX yap5Δ::URA3/YAP5-Malaysian</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>YHL410</td>
<td>YHL243 MATα/α ura3Δ:kanMX hoΔ:hpMX yap5Δ::URA3/YAP5-Malaysian</td>
<td>this study</td>
<td></td>
</tr>
<tr>
<td>STRAIN NAME</td>
<td>PARENTS</td>
<td>GENOTYPE</td>
<td>SOURCE</td>
</tr>
<tr>
<td>-------------</td>
<td>---------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>YHL411</td>
<td>YHL243</td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX yap5Δ::URA3/YAP5- WineEuropean</td>
<td>this study</td>
</tr>
<tr>
<td>YHL413</td>
<td>YHL247</td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX yap5Δ::URA3/YAP5- Malaysian</td>
<td>this study</td>
</tr>
<tr>
<td>YHL414</td>
<td>YHL247</td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX yap5Δ::URA3/YAP5- WineEuropean</td>
<td>this study</td>
</tr>
<tr>
<td>YHL415</td>
<td>YHL247</td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX yap5Δ::URA3/YAP5- Malaysian</td>
<td>this study</td>
</tr>
<tr>
<td>YHL449</td>
<td>YHL243</td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX ccc1Δ::URA3/CCC1- WineEuropean</td>
<td>this study</td>
</tr>
<tr>
<td>YHL450</td>
<td>YHL243</td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX ccc1Δ::URA3/CCC1- Malaysian</td>
<td>this study</td>
</tr>
<tr>
<td>YHL451</td>
<td>YHL243</td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX ccc1Δ::URA3/CCC1- Malaysian</td>
<td>this study</td>
</tr>
<tr>
<td>YHL452</td>
<td>YHL247</td>
<td>MATa/α ura3Δ::kanMX hoΔ::hphMX ccc1Δ::URA3/CCC1- WineEuropean</td>
<td>this study</td>
</tr>
</tbody>
</table>
Supplementary Table 3.S2 — RNAseq statistics.

<table>
<thead>
<tr>
<th>LIBRARY</th>
<th>TOTAL READS</th>
<th>UNIQUELY MAPPING READS</th>
</tr>
</thead>
<tbody>
<tr>
<td>UWOPS03.461.4</td>
<td>70,275,289</td>
<td>37,992,498</td>
</tr>
<tr>
<td>UWOPS05.217.3</td>
<td>76,341,920</td>
<td>34,547,610</td>
</tr>
<tr>
<td>UWOPS05.227.2</td>
<td>63,801,895</td>
<td>28,127,756</td>
</tr>
<tr>
<td>BC187</td>
<td>62,813,354</td>
<td>35,442,506</td>
</tr>
<tr>
<td>RM11-1</td>
<td>76,501,254</td>
<td>35,055,758</td>
</tr>
<tr>
<td>UWOPS03.461.4 x BC187</td>
<td>137,307,729</td>
<td>28,763,569</td>
</tr>
<tr>
<td>UWOPS03.461.4 x RM11-1</td>
<td>151,386,079</td>
<td>29,704,996</td>
</tr>
<tr>
<td>UWOPS03.461.4 x RM11-1</td>
<td>164,905,032</td>
<td>31,783,714</td>
</tr>
<tr>
<td>UWOPS05.217.3 x BC187</td>
<td>47,399,026</td>
<td>6,133,068</td>
</tr>
</tbody>
</table>
Table 3.1 — Directional expression divergence between Malaysian and wine-European yeast in co-regulated gene groups. Each row reports the results of a test for expression changes of the same sign between Malaysian and wine/European strains in one group of functionally related genes. Upregulated, population with elevated expression. Adjusted $p$-value, significance of a two-sided resampling test relative to the genomic null for an extreme value of the sum, across genes of the indicated regulon, of the signs of the log$_2$ ratio of expression in Malaysian yeast to wine/European yeast, corrected for multiple testing with the Benjamini-Hochberg method. Shown is the set of tested groups significant at adjusted $p < 0.05$.

<table>
<thead>
<tr>
<th>GROUP</th>
<th>ANNOTATION</th>
<th>UPREGULATED</th>
<th>ADJUSTED $P$-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node_7</td>
<td>Ribosomal proteins</td>
<td>Malaysian</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>$AFT1$ targets</td>
<td>Iron homeostasis</td>
<td>Wine/European</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Node_36</td>
<td>Ribosomal proteins</td>
<td>Malaysian</td>
<td>0.0002</td>
</tr>
<tr>
<td>Lysine cluster</td>
<td>Lysine utilization</td>
<td>Wine/European</td>
<td>0.0002</td>
</tr>
<tr>
<td>Node_48</td>
<td>Ribosomal proteins</td>
<td>Wine/European</td>
<td>0.0002</td>
</tr>
<tr>
<td>$GCN4$ chIP</td>
<td>Amino acid metabolism</td>
<td>Wine/European</td>
<td>0.0004</td>
</tr>
<tr>
<td>GO:0055085</td>
<td>Transmembrane transport</td>
<td>Wine/European</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>GO:0007039</td>
<td>Vacuolar protein catabolic process</td>
<td>Wine/European</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>GO:0006519</td>
<td>Cellular amino acid metabolic process</td>
<td>Wine/European</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>GO:0006725</td>
<td>Cellular aromatic acid metabolic process</td>
<td>Wine/European</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>GO:0006826</td>
<td>Iron ion transport</td>
<td>Wine/European</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>GO:0006812</td>
<td>Cation transport</td>
<td>Wine/European</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>GO:0006811</td>
<td>Ion transport</td>
<td>Wine/European</td>
<td>0.0188</td>
</tr>
<tr>
<td>GO:0006082</td>
<td>Organic acid metabolic process</td>
<td>Wine/European</td>
<td>0.0329</td>
</tr>
</tbody>
</table>
Figure 3.1 — Directional changes in regulon expression between Malaysian and wine/European yeast. In (A)-(C), each panel shows expression divergence between Malaysian and wine/European yeast in one suite of iron-metabolism genes. Each cell reports the expression divergence of one gene measured as a log₂ ratio. Each column shows data from a comparison between one Malaysian isolate and one Wine/European isolate (M1, UWOPS03.461.4; M2, UWOPS05.217.3; M3, UWOPS05.227.2; WE1,
RM11-1; WE2, BC187); each row shows data for one gene. (A) Gene targets of the AFT1 transcription factor, from (Gasch et al., 2004). (B) Genes in the Gene Ontology term GO:0006826, iron ion transport. (C) Genes upregulated in a Δccc1 laboratory strain compared to wild-type during exposure to high-iron medium (Lin et al., 2011). (D) Cartoon of iron homeostasis regulation in laboratory yeast. Under conditions of iron scarcity (left) the transcription factor AFT1 upregulates metal transporters and other iron uptake factors. Under conditions of excess iron (right), iron uptake genes are downregulated, and the transcription factor YAP5 upregulates iron resistance genes, including the vacuolar transporter CCC1 and other genes functioning in function in iron storage, iron-sulfur cluster enzymes, respiration, DNA damage repair and oxidative stress responses.
Supplementary Figure 3.S1 — Lack of allele-specific expression in iron homeostasis genes. Each panel shows measurements of allele-specific expression in Malaysian-wine/European hybrids in one suite of iron-metabolism genes. Each cell reports the log$_2$ ratio of allele-specific expression levels of the Malaysian and wine/European alleles of one gene in a hybrid between a Malaysian and wine/European strain, measured using reads uniquely mapped to each parent’s allele in turn. Each column reports measurements from one hybrid (M1, UWOPS03.461.4; M2, UWOPS05.217.3; WE1, RM11-1; WE2, BC187), and each row shows data for one gene. (A) Gene targets of the AFT1 transcription factor, from (Gasch et al., 2004). (B) Genes in the Gene Ontology term GO:0006826, iron ion transport. (C) Genes upregulated in a Δccc1 laboratory strain compared to wild-type during exposure to high-iron medium (Lin et al., 2011).
Shown are protein-coding

\[ \text{Sequence variation across yeast strains in genes} \]

Supplementary Figure 3.S2 — Sequence variation across yeast strains in genes underlying iron susceptibility in Malaysian isolates. Shown are protein-coding
alignments using sequence and strain identifiers from (Liti et al., 2009) for regions of (A) \textit{YAP5} and (B) \textit{CCC1}. Non-synonymous coding variants private to the Malaysian population are indicated in red.
Figure 3.2 — **YAP5 and CCC1 underlie growth defects of Malaysian yeast in high-iron conditions.** Each bar reports the mean doubling time of one yeast strain during log-phase growth in the indicated condition. In each panel, bars at left report growth of parent strains and wild-type hybrids, with Malaysian and Wine/European indicating the average growth over all Malaysian strains (UWOPS03.461.4, UWOPS05.217.3 and UWOPS05.227.2) and Wine/European strains (RM11-1 and BC187), respectively, and Malaysian x Wine/European indicating growth of the UWOPS03.461.4 x BC187 hybrid. Center bars report growth of the **YAP5** reciprocal hemizygotes in the UWOPS03.461.4 x BC187 hybrid background with yap5-YAP5-Malaysian indicating the strain harboring only the allele of **YAP5** from the Malaysian strain, and yap5/YAP5-Wine/European indicating the strain with the Wine/European allele of **YAP5**. Bars at right report growth of **CCC1** reciprocal hemizygotes in the UWOPS03.461.4 x BC187 hybrid background, with ccc1/CCC1-Malaysian indicating the strain harboring only the allele of **CCC1** from the Malaysian strain and ccc1/CCC1-Wine/European indicating the strain with the Wine/European allele of **CCC1**. (A) Complete media without FeSO$_4$ and (B) complete media with 5 mM FeSO$_4$. 

37
Table 3.2 — Elevated protein evolutionary rates in iron-toxicity response genes in Malaysian yeast. Each row reports analysis of the protein evolutionary rate of the set of genes upregulated in a Δccc1 laboratory strain compared to wild-type during exposure to high-iron conditions from (Lin et al., 2011). Each p-value reports the significance of a one-sided resampling test relative to the genomic null for a greater value of the mean, across genes of the iron-toxicity response regulon, of ratios of non-synonymous to synonymous coding sequence variants private and fixed to the indicated yeast population.

<table>
<thead>
<tr>
<th>POPULATION</th>
<th>REGULON MEAN $K_a/K_s$</th>
<th>GENOME MEAN $K_a/K_s$</th>
<th>P-VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wine/European</td>
<td>0.1504</td>
<td>0.1240</td>
<td>0.2530</td>
</tr>
<tr>
<td>Malaysian</td>
<td>0.2629</td>
<td>0.1649</td>
<td>0.0435</td>
</tr>
<tr>
<td>Sake</td>
<td>0.0594</td>
<td>0.1014</td>
<td>0.7480</td>
</tr>
<tr>
<td>North American</td>
<td>0.0455</td>
<td>0.0821</td>
<td>0.7751</td>
</tr>
<tr>
<td>West African</td>
<td>0.2204</td>
<td>0.1777</td>
<td>0.2080</td>
</tr>
</tbody>
</table>
REFERENCES


