Divergence of iron metabolism in wild Malaysian yeast

Hana N. Lee‡, Yulia Mostovoy‡, Tiffany Y. Hsu§, Amanda H. Chang, and Rachel B. Brem*

Department of Molecular and Cell Biology, University of California, Berkeley
‡Current address: Department of Ecology and Evolution, University of Chicago, Chicago, IL
§Current address: Graduate Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA
*These authors contributed equally to this work.
*To whom correspondence should be addressed: rbrem@berkeley.edu

Keywords

Iron metabolism, yeast, regulatory variation
Abstract

Comparative-genomic studies have reported widespread variation in levels of gene expression within and between species. Using these data to infer organism-level trait divergence has proven to be a key challenge in the field. We have used a wild Malaysian population of *S. cerevisiae* as a testbed in the search to predict and validate trait differences based on observations of regulatory variation. Malaysian yeast, when cultured in standard medium, activated regulatory programs that protect cells from the toxic effects of high iron. Malaysian yeast also showed a hyperactive regulatory response during culture in the presence of excess iron, and had a unique growth defect in high-iron conditions. Molecular validation experiments pinpointed the iron metabolism factors *AFT1*, *CCC1*, and *YAP5* as contributors to these molecular and cellular phenotypes, and in genome-scale sequence analyses, a suite of iron-toxicity response genes showed evidence for rapid protein evolution in Malaysian yeast. Our findings support a model in which iron metabolism has diverged in Malaysian yeast as a consequence of a change in selective pressure, with Malaysian alleles shifting the dynamic range of iron response to low iron concentrations and weakening resistance to extreme iron toxicity. By dissecting the iron-scarcity specialist behavior of Malaysian yeast, our work highlights the power of expression divergence as a signpost for biologically and evolutionarily relevant variation at the organismal level.
Article summary

Interpreting the phenotypic relevance of gene expression variation is one of the primary challenges of modern genomics. Observing striking changes in iron metabolism gene expression between wild yeast populations, we hypothesized that these microbes would differ with respect to growth during iron challenge. Fitness measurements validated this prediction, revealing a growth defect in wild Malaysian yeast under conditions of high iron; molecular-genetic dissection identified three genes that contribute to molecular and cellular divergence in iron metabolism between yeast strains. Our results make clear that expression profiles can be used as a powerful hypothesis generator for the study of trait variation.
Introduction

Heritable genetic differences at the molecular level are widespread between populations and species. Much of the divergence observed in DNA sequence, gene expression, and other molecular attributes may bear little relationship to macroscopic traits, and the outstanding challenge in the field is to detect signatures of biological relevance in catalogs of molecular data. This unbiased search for trait differences with an impact on fitness, motivated by screens of molecular observations alone, is often referred to as reverse ecology (Li et al. 2008b).

The reverse ecology approach is particularly pertinent to the study of regulatory divergence, given that comparisons between species routinely detect expression change at more than half of the genes in a given genome (Wilson and Odom 2009; Dowell 2010). Although linkage or association methods can map the determinants of evolutionarily relevant trait differences within species to loci that affect gene expression (Shapiro et al. 2004; Rebeiz et al. 2009; Tung et al. 2009; Wittkopp et al. 2009; Loehlin and Werren 2012), these methods cannot be applied to comparisons between reproductively isolated populations. In landmark cases, focused candidate-gene studies have implicated regulatory changes in trait differences between species (Marcellini and Simpson 2006; Hanikenne et al. 2008; Booth et al. 2010; Werner et al. 2010). The ultimate goal is to predict and test phenotypes ab initio from genome-scale expression profiles of divergent individuals; to date, little precedent has been set for the experimental validation of such a pipeline.

In this work, we set out to use wild yeast as a model system for the reverse ecology paradigm. Nectar of the flowers of Malaysian bertam palm trees hosts a yeast community whose fermentation is consumed by mammalian pollinators (Wiens et al. 2008). S. cerevisiae strains collected from this nectar are reproductively isolated from the rest of the species (Cubillos et al. 2011), so approaches other than standard mapping methods are required to investigate the evolution of this unique yeast population. An initial, unbiased transcriptional profiling experiment led us to the discovery of divergent iron-homeostasis and iron-sensitivity behaviors in the Malaysian strains, and to a dissection of the genetic and evolutionary basis of these phenotypes.

Results

To survey gene expression programs in wild yeast, we took a comparative approach using three homozygous Malaysian isolates and two homozygotes from a distinct, well-defined wine/European yeast population (Lití et al. 2009). We cultured each strain in rich medium and subjected each to transcriptional profiling by RNA-seq (Supplementary Table 1). We used the RNA-seq reads to verify and correct coding sequences from low-coverage genomes available for these strains (Lití et al. 2009), identifying a total of 46,367 single-nucleotide polymorphisms in coding regions in which all Malaysian strains sequenced harbored one allele, and all wine/European strains harbored a second allele (Supplementary Data Set 1). Mapping to the amended genomes and analysis of the
complete set of profiles revealed 601 genes differentially expressed between the populations at a false discovery rate (FDR) of 5% (Supplementary Data Set 2).

We sought to focus on patterns of regulatory variation across groups of functionally related genes, which could be a signpost for organismal trait differences between Malaysian and wine/European yeast. For this purpose, we screened gene groups for coherent expression change between the populations. Testing groups delineated by Gene Ontology (GO; http://yeastgenome.org) and those defined on the basis of co-regulation in classic analyses of laboratory yeast (Gasch et al. 2004), we identified seven GO terms and three regulons whose genes were predominantly upregulated, or predominantly downregulated, in Malaysian strains relative to wine/European strains (FDR = 5%; Table 1). Among these were two sets of genes annotated in iron metabolism, for which the Malaysian population was associated with low expression (Figure 1A, B): targets of Aft1p, a transcription factor regulating iron-uptake genes, and the iron ion transport Gene Ontology term GO:0006826. This largely non-overlapping pair of gene groups comprises genes which, in studies of laboratory yeast, are expressed at low levels in the presence of excess iron, moderate levels in standard medium, and high levels during iron starvation (Yamaguchi-Iwai et al. 1996; Yamaguchi-Iwai et al. 2002; Shakoury-Elizeh et al. 2004) (Figure 1D).

Given the divergent expression of iron-starvation genes between Malaysian and wine/European yeasts, we hypothesized that regulation of factors that protect against the toxic effects of excess iron might also be distinct between these two populations. To test this, we used a previously defined suite of genes upregulated by laboratory yeast in iron-toxic conditions (Lin et al. 2011), which includes components of the DNA damage and oxidative stress response pathways that we refer to as iron-resistance genes. Although, as expected, in the absence of excess iron this regulon was not highly expressed in either population, mRNA levels were modestly but significantly elevated in the Malaysian population (Figure 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 1D).

We hypothesized that the molecular differences in iron metabolism between yeast populations could have organism-level correlates in the context of growth during iron challenge. To test this, we first measured growth rates of Malaysian and wine/European parent strains, as well as the wild-type hybrid formed by a mating between them, in iron-toxicity conditions and in a complete-medium control. In the presence of excess iron, Malaysian parent strains had a marked growth defect, with doubling times up to 1.7-fold elevated relative to wine/European isolates (Figure 2 and Supplementary Data Set 3). 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 2 and Supplementary Data Set 3). We also assessed growth of other environmental yeast isolates and found all to be resistant to iron toxicity (Figure 3 and Supplementary Data Set 4), with the exception ofNCYC110, a member of the broadly stress-sensitive West African population (Zorgo et al. 2012).
We next sought to dissect the genetic basis of divergence in iron metabolism regulation among Malaysian strains using a candidate-gene framework. We expected that causative variants would likely act in \textit{trans} on the iron metabolism machinery, since measurements of allele-specific expression in Malaysian x wine/European hybrids showed no strong signals of coherent \textit{cis}-regulatory variation in iron metabolism genes (Supplementary Table 1 and Supplementary Data Sets 5 and 6). Likewise, sequence-based analyses revealed no striking promoter variation in iron metabolism genes (data not shown). To identify candidate \textit{trans}-acting loci underlying growth and regulatory traits, 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 \textit{et al.} 2001; Li \textit{et al.} 2008a): the transcription factor \textit{YAP5}, which activates genes in response to treatment with excess iron, and \textit{CCC1}, a transporter that sequesters iron ions in the vacuole (Figure 1D). We also chose for molecular validation the transcription factor \textit{AFT1}, a protein which has been characterized for its activity in low levels of iron as a regulator of the iron-scavenging genes emerging from our original RNA-seq analysis (Shakoury-Elizeh \textit{et al.} 2004) (Figure 1A and D). In all three protein sequences, we observed non-conservative amino acid changes in Malaysian strain genomes (Supplementary Figure 1).

To investigate the role of our candidate genes in the Malaysian growth phenotype in high iron, we used reciprocal hemizygote analysis (Steinmetz \textit{et al.} 2002) for each candidate locus 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 hybrid strains in such a pair 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 cultured reciprocal hemizygote strains for \textit{AFT1}, \textit{YAP5}, and \textit{CCC1} in high-iron conditions and in a complete-medium control, alongside the wild-type parents, and we measured growth rates of each strain in each case. The results revealed a 1.2-fold difference in iron resistance between strains bearing the Malaysian and wine/European alleles of \textit{CCC1}, and a more modest (1.05-fold), though still significant, effect on iron resistance of variation in \textit{YAP5} (Figure 2 and Supplementary Data Set 3). In each case, the Malaysian allele of the respective gene compromised growth in high-iron conditions, mirroring the known effects of engineered null mutations at these loci (Li \textit{et al.} 2001; Li \textit{et al.} 2008a).

Expression measurements also supported a model of losses of function at Malaysian orthologs of \textit{CCC1} and \textit{YAP5}: the Malaysian allele of \textit{CCC1} drove upregulation of iron-starvation genes and iron-resistance genes in standard complete medium and high-iron conditions, respectively, while the Malaysian allele of \textit{YAP5} reduced expression of iron-resistance genes in high-iron conditions (Figure 4, Supplementary Figure 2, and Supplementary Data Sets 7 and 8), consistent with the behavior of hypomorphs at these loci (Li \textit{et al.} 2001; Li \textit{et al.} 2008a; Lin \textit{et al.} 2011). Additionally, strains hemizygous for \textit{CCC1} grew more slowly than wild-type hybrids in high-iron medium regardless of which
allele they carried (Figure 2 and Supplementary Data Set 3), reflecting haploinsufficiency of \textit{CCC1} with respect to iron resistance. As expected, variation in \textit{AFT1}, which functions primarily in conditions of iron starvation (Figure 1D), had little impact on growth when iron was present in excess (Figure 2 and Supplementary Data Set 3). However, the Malaysian allele of \textit{AFT1} had a significant regulatory effect in reciprocal hemizygotes, driving upregulation of iron-starvation genes during growth in complete medium (Figure 4, Supplementary Figure 2, and Supplementary Data Set 7) in a manner consistent with a gain-of-function mutation (SHAKOURY-ELIZEH \textit{et al.} 2004). We conclude that the growth defect of Malaysian yeast in the presence of excess iron is due in part to alleles that act at losses of function at the iron-resistance genes \textit{YAP5} and \textit{CCC1}. With respect to gene regulation, the Malaysian genome harbors alleles with both activating and repressing effects on iron metabolism, reflecting a complex genetic model that involves \textit{YAP5} and \textit{CCC1} as well as \textit{AFT1}.

We next aimed to investigate signatures of natural selection on iron metabolism in Malaysian yeast, complementing our expression- and phenotype-based analyses with sequence-based tests. Using single-nucleotide polymorphisms private to and conserved in each of five yeast populations characterized by (LITI \textit{et al.} 2009), we estimated ratios of non-synonymous to synonymous substitution rates in coding regions of iron-metabolism genes and compared these rates to those of the rest of the genome. The results, shown in Table 2, revealed a 1.6-fold elevation of protein evolutionary rate in Malaysian yeast in genes activated during high-iron exposure, with no such effect detectable in any other yeast population. Iron-starvation genes exhibited no signature of non-neutral sequence divergence in any population (data not shown). These results provide evidence for a unique history of a change of selective pressure on iron-resistance genes 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, the power of expression divergence as a predictor of trait variation has remained largely unvalidated. To establish Malaysian yeast as a testbed for the reverse ecology approach, we generated expression profiles from cultures grown in standard conditions, and we used these data as a point of departure for a study of regulatory and growth phenotypes in the presence of excess iron. Though several elegant studies have characterized growth traits in yeast populations (WARRINGER \textit{et al.} 2011; HODGINS-DAVIS \textit{et al.} 2012; ZORGO \textit{et al.} 2012), to date the degree of divergence between strains in iron sensitivity has been unknown. Our expression measurements, growth assays, and molecular-genetic manipulations confirmed the iron sensitivity of Malaysian yeast and the role of three candidate genes, \textit{AFT1}, \textit{YAP5}, and \textit{CCC1}, in regulatory and growth phenotypes.

What are the evolutionary forces that have driven divergence of iron metabolism in Malaysian yeast? The behavior of the Malaysian \textit{YAP5} and \textit{CCC1} alleles as losses of function, and the sequence-based evidence for non-neutral evolution of the coding
regions of iron-resistance genes, support a model of specialization by Malaysian strains to a low-iron environment. In one evolutionary scenario, reduced exposure to high-iron conditions in the Malaysian niche would have relaxed the strength of purifying selection on the iron-toxicity response, allowing Malaysian yeasts to accumulate mutations in this gene network. Additionally, given the gain-of-function behavior of the Malaysian allele of AFT1, it is tempting to speculate that negative regulatory control of this activator of the iron-starvation response has been eliminated in Malaysian yeast, to sidestep iron-sensing mechanisms and raise expression of iron transporters and scavengers as would be suitable for a constant, low-iron environment. By contrast, the avid iron uptake and functional vacuolar iron storage of wine/European strains would reflect a need for flexible iron-homeostasis machinery that can respond more fully to environmental change. Taken together, our findings provide a case in which the rare and often deleterious mutations that litter wild yeast genomes (ZORGO 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 microbes as a driver of evolutionary change.

Although we successfully harnessed expression profiles to identify an evolutionarily relevant trait difference between yeast populations, our results leave open the question of the molecular basis of the expression patterns themselves. In reciprocal hemizygote experiments, the most striking regulatory effects of our candidate genes were in directions that opposed the patterns of divergence in parental homozygotes: Malaysian alleles at the YAP5 and AFT1 transcription factors drove downregulation of iron-resistance genes and upregulation of iron-starvation genes, respectively, contrasting with the high and low expression of these gene sets, respectively, in Malaysian parental strains. Given the prevalence of transgressive segregation in the genetics of gene expression variation within species (GIBSON and WEIR 2005), we favor a model in which the regulatory program of Malaysian homozygotes is an indirect regulatory response to a suite of defects in iron-resistance genes in Malaysian homozygotes, one which obscures the transgressive contribution of YAP5 and AFT1 in the purebred context. However, our experiments do not rule out the possibility that alleles of YAP5 and AFT1 exert different effects in the parental background than they do in the hybrid. Future work will establish the potential for epistasis among our candidate loci and between each gene and the genetic background, using the loci we have validated here as a springboard to deepen our mechanistic understanding of the regulatory and growth behaviors of Malaysian isolates.

Our work makes clear that expression profiles can be used as a powerful hypothesis generator for the study of organismal trait variation. This reverse ecology paradigm is likely to be most successful when coherent regulatory change in well-annotated pathways underlies, or responds to, change in a phenotype. With the increased availability of functional-genomic resources in many taxa, the expression-based approach will likely prove to be broadly applicable in the genetic dissection of differences between populations and species.
Materials and Methods

RNA-seq library preparation

All strains used are listed in Supplementary Table 2. 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. For analysis of cis-regulatory variation in Supplementary Data Set 3, three Malaysian x wine/European hybrids were generated through single-cell matings between Malaysian and wine/European strains: UWOPS03.461.4 x RM11-1, UWOPS03.461.4 x BC187 and UWOPS05.217.3 x BC187.

Parental and hybrid strains were grown to mid-log phase (0.65-0.72 OD_{600}) in yeast peptone dextrose (YPD) medium (AMBERG et al. 2005) at 30°C. Total RNA was extracted from cells using the hot phenol method (AUBERG et al. 2005), 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 base-pair 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) using Bowtie (LANGMEAD et al. 2009). Single-nucleotide polymorphisms (SNPs) were called using SAMtools (LI et al. 2009) and, for those with the maximum genotype quality Phred score of 99, were used to amend the genome sequences downloaded from (LITI et al. 2009). These amended genomes were then used as references for re-mapping of RNA-seq reads, with only the reads mapping uniquely and containing no mismatches used for further analysis. For each Malaysian x wine/European hybrid, RNA-seq reads were mapped as above to a genome file formed by concatenation of the amended genomes of the respective parents, yielding only reads that mapped uniquely to regions containing polymorphisms between the Malaysian and wine/European parents and thus reported allele-specific expression.

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 to correct for differences in library size between lanes and the loess method to correct for differences in GC content within lanes. For analysis of differential expression between wine/European and Malaysian yeast in Table 1 and Supplementary Data Set 2, for a given gene the multiple isolates from each population were used as replicates in the DESeq package (ANDERS and HUBER 2010) to calculate a composite ratio of the expression levels in the two populations and for assessment of statistical significance, with multiple-testing correction across genes using the Benjamini-Hochberg method. We used the same approach for analysis of allele-specific expression in Supplementary Data Set 5: for a given gene, read counts per allele from each of the three hybrids were
used as input into DESeq to calculate a composite gene expression ratio and significance of the difference between the alleles. As a point of comparison for the calculations of differential expression between parental species and between alleles in the hybrid, we repeated the read-mapping pipeline as above using un-corrected genome sequences from (LITI et al. 2009), and re-applied tests for expression change to the estimates from this procedure. At corrected \( p < 0.05 \) these calculations identified 480 and 0 genes differentially expressed between parents and alleles in the hybrid, respectively (data not shown), in contrast to the 601 and 265, respectively, reported at this significance level in Supplementary Data Sets 2 and 5.

**Analysis of directional regulatory divergence in gene groups**

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; 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 taking the log\(_2\) fold change between the populations for each such gene as described above, we summed these differential expression measures across the genes of the regulon. We then repeated this ratio 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 ratio 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 \texttt{p.adjust()} function in R. For analysis of directional \textit{cis}-regulatory variation in Supplementary Data Set 6, we repeated this group test procedure using as input, for a given gene, the log\(_2\) ratio of the expression of the Malaysian allele to that of the wine/European allele, measured in hybrids as described above.

To analyze iron-toxicity response genes in Figure 1 and Table 2, we used the set of genes upregulated in a \textit{ccc1} laboratory strain, compared to wild-type, in 3 mM FeSO\(_4\) from (LIN et al. 2011). Expression analysis was as described above and sequence analysis as below. For analysis of iron-toxicity response factors in Figure 4 and Supplementary Figure 2, we also mined the set of genes responsive to 2 mM FeSO\(_4\) in (PIMENTEL et al. 2012).

**Reciprocal hemizygote strain construction**

Malaysian x wine/European hybrid strains (YHL243 and YHL247, both UWOPS03.461.4 x BC187 hybrids from reciprocal matings; see Supplementary Table 2) were used to make reciprocal hemizygotes by deleting one allele of the gene of interest (\textit{CCC1}, \textit{YAP5}, or \textit{AFT1}) with a \textit{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.
Growth assays

For each strain, a pre-culture was prepared by growth for four hours in YPD medium and diluted to 0.1 OD\textsubscript{600} by resuspension in distilled water. Experimental cultures were then established in triplicate for each strain by diluting this pre-culture 1:10 in complete synthetic medium (CSM) (Amberg \textit{et al.} 2005) supplemented with 5 mM FeSO\textsubscript{4}, or supplemented with an equal amount of water as an untreated control, for a starting concentration of 0.01 OD\textsubscript{600} in a 96-well plate to assess growth. Growth was monitored via OD\textsubscript{600} measurements every 30 minutes in a Tecan GENios microplate reader. Growth attributes in Figure 2 and Supplementary Data Sets 3 and 4 were calculated from growth curves using the method of (Warringer \textit{et al.} 2011) implemented in Python. In Figure 2 and Supplementary Data Set 3, culturing and analysis were carried out for 2-4 transformants of each reciprocal hemizygote.

Quantitative PCR

Two biological replicates of each parental strain, two biological replicates of each of two independent transformants of each AFT1 and YAP5 reciprocal hemizygote, and two sets of CCC1 reciprocal hemizygote strains, each comprising two biological replicates of each of two independent transformants, were grown to mid-log phase in YPD medium at 30°C. For iron-treated samples, FeSO\textsubscript{4} was added to a final concentration of 5 mM. After 30 minutes of incubation at 30°C, total RNA was isolated and treated with DNase I as described above for RNA-seq, except that for iron-treated samples, RNA was purified on RNeasy columns (Qiagen). For each sample, cDNA was synthesized using SuperScript III reverse transcriptase (Life Technologies) and diluted to 1 ng/µL. Maxima SYBR green (Thermo Scientific) was used for quantitative PCR, with 2-3 technical replicates for each gene in each strain, on a Mx3000P system (Agilent). Cycle thresholds (Ct) for each replicate were obtained using the MxPro software (Agilent), and gene expression levels relative to ACT1 were calculated using the 2\textsuperscript{ΔΔCt} method (Livak and Schmittgen 2001). As a further normalization step for ease of comparison between experiments, the vector of expression measures of a given gene in a given strain background and replicate set was centered with respect to the median across the set. Statistical significance of differential expression between genotypes for a given regulon was assessed by a paired Wilcoxon test. YAP5 reciprocal hemizygotes did not differ significantly with respect to iron-starvation gene expression, and we detected little impact of variation at AFT1, YAP5, or CCC1 on iron-resistance genes in synthetic complete medium (data not shown).

K\textsubscript{a}/K\textsubscript{s} 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 \textit{et al.} 2009) using uncorrected published sequences. The ratio of nonsynonymous to synonymous substitution rates, using the formula for K\textsubscript{a}/K\textsubscript{s} in (Nei and Gojobori 1986), was calculated for each gene using SNPs fixed in and private to each non-mosaic yeast population, with populations as defined in (Liti \textit{et al.} 2009). For each population, we...
calculated the mean of these $K_a/K_s$ values across all genes of the iron-toxicity response regulon as defined above. We then repeated this mean $K_a/K_s$ calculation 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 mean as extreme as, or more extreme than, the mean of the real regulon was taken to be the empirical $p$-value for protein evolutionary rate.

Acknowledgements

We thank Jeremy Roop, Christopher Nelson, and Alexander Thiemecke for molecular biology expertise, and Jasper Rine for his generosity with advice and resources. This work was supported by an Ellison New Scholar Award in Aging, NIH R01 GM087432, NIH R21 NS077224, and NIH R01 AG043080 to RBB, and a Paul and Daisy Soros Fellowship for New Americans to HNL.

Authors’ contributions

HNL carried out all data analysis, all strain construction, and all growth assays, and wrote the manuscript. TYH constructed RNA-seq libraries. YM and AC carried out QPCR assays. RBB conceived the study design and wrote the manuscript. All authors read and approved the final manuscript.
Figure captions

Figure 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, with the complete set of genes from a previously defined regulon shown in each case. Each cell reports the log₂ of the expression ratio of one gene between two strains: 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), and each row shows data for one gene. (A) Gene targets of the Aft1p transcription factor, from (GASCH et al. 2004); directional differential expression between populations significant at corrected \( p = 0.0078 \) (Table 1). (B) Genes in the Gene Ontology term GO:0006826, iron ion transport; directional expression \( p = 0.04 \) (Table 1). (C) Genes upregulated in a ccc1 laboratory strain compared to wild-type during exposure to high-iron medium (LIN et al. 2011); directional expression \( p = 0.011 \) (Table 1). (D) Cartoon of iron homeostasis regulation in laboratory yeast. Under conditions of iron scarcity (left) the transcription factor Aft1p upregulates iron-uptake and iron-conservation factors. Under conditions of excess iron (right), iron-starvation genes are downregulated, and the transcription factor Yap5p upregulates iron-toxicity response factors, including the vacuolar transporter CCC1 and other genes functioning in iron storage, iron-sulfur cluster enzyme biogenesis, respiration, DNA damage repair and oxidative stress response.

Figure 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, Malaysian and wine/European indicate the average growth across biological replicates of the Malaysian strain UWOPS03.461.4 \((n = 9)\) and the wine/European strain BC187 \((n = 9)\), respectively. Malaysian x wine/European indicates growth of the UWOPS03.461.4 x BC187 hybrid \((n = 15)\). The fourth and fifth bars of each panel report growth of YAP5 reciprocal hemizygotes in the UWOPS03.461.4 x BC187 hybrid background, with yap5/YAP5-Malaysian denoting strains harboring only the allele of YAP5 from the Malaysian strain \((n = 12)\), and yap5/YAP5-wine/European denoting the strains with the wine/European allele of YAP5 \((n = 6)\). The sixth and seventh bars report growth of CCC1 reciprocal hemizygotes in the hybrid background, with ccc1/CCC1-Malaysian denoting hemizygotes harboring only the allele of CCC1 from the Malaysian strain \((n = 12)\), and ccc1/CCC1-wine/European denoting hemizygotes with the wine/European allele of CCC1 \((n = 6)\). The last two bars report growth of AFT1 reciprocal hemizygotes in the hybrid background, with aft1/AFT1-Malaysian denoting hemizygotes harboring only the allele of AFT1 from the Malaysian strain \((n = 6)\), and aft1/AFT1-wine/European denoting hemizygotes with the wine/European allele of AFT1 \((n = 6)\). Error bars in all panels represent 95% confidence intervals. (A) Complete media without additional FeSO₄. (B) Complete media supplemented with 5 mM FeSO₄. ***, significant at Wilcoxon \( p < 0.0005 \); **, Wilcoxon \( p < 0.005 \); *, Wilcoxon \( p < 0.05 \). Raw measurements are reported in Supplementary Data Set 3.
Figure 3. Most non-Malaysian yeast isolates are resistant to high-iron toxicity. Each bar reports the mean doubling time of one yeast strain ($n = 3$) during log-phase growth in the indicated condition. Each color reports growth of strains from one yeast population as defined in (Litì et al. 2009): blue, Malaysian (from left to right, strains UWOPS03.461.4, UWOPS05.217.3, UWOPS05.227.2); red, wine/European (BC187, RM11-1); green, North American (YPS606, YPS128); orange, West African (NCYC110, DBVP6044); yellow, sake (Y12); grey, laboratory (S288C). (A) Complete media without FeSO$_4$. (B) Complete media with 5 mM FeSO$_4$. *** significant at Wilcoxon $p < 0.0005$. Error bars represent 95% confidence intervals. Raw measurements are reported in Supplementary Data Set 4.

Figure 4. Regulatory impact of variation in $AFT1$, $CCC1$, and $YAP5$ between Malaysian and wine/European yeast. Each panel reports the regulatory effects of variation at iron-metabolism genes in yeast cultured in one growth medium. In a given panel, each row reports comparisons between the effects of Malaysian and wine/European genotypes on expression of an iron-starvation (Gasch et al. 2004) or iron-resistance (Lin et al. 2011; Pimentel et al. 2012) target gene. Color in each cell represents a ratio of the expression measurements from two strains. A given cell in the first column reports the ratio, for the indicated target gene, of the expression in a homozygous Malaysian strain (UWOPS03.461.4), as a median across replicates ($n = 2$), to that in a homozygous wine/European strain (BC187). A given cell in each remaining column reports expression as a median across replicates, in a Malaysian-wine/European hemizygote (UWOPS03.461.4 x BC187) bearing the Malaysian allele of the indicated variant locus ($CCC1$, $n = 8$; $AFT1$, $n = 4$; $YAP5$, $n = 4$) relative to the median expression in the hemizygote bearing the wine/European allele. (A) Iron-starvation genes in synthetic complete medium. (B) Iron-resistance genes in synthetic complete medium supplemented with 5 mM FeSO$_4$. Values at the top of each column report the results of a paired Wilcoxon test for the significance of the differences in expression, across genes of the regulon, between the indicated genotypes. Individual replicate measurements are visualized in Supplementary Figure 2 and raw data are reported in Supplementary Data Sets 7 and 8.
Table captions

**Table 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. Group, identifier of regulon from (GASCH et al. 2004) or Gene Ontology term. 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, assessed using all isolates from each population and corrected for multiple testing with the Benjamini-Hochberg method. Shown is the set of tested groups significant at adjusted $p < 0.05$.

**Table 2. Elevated protein evolutionary rates in iron-toxicity response genes in Malaysian yeast.** Each row reports analysis of the protein evolutionary rate, in one yeast population, of the set of genes upregulated in a ccc1 laboratory strain compared to wild-type during exposure to high-iron medium (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 to and fixed in the indicated yeast population.
Supplementary figure captions

Supplementary Figure 1. Coding variants private to Malaysian yeast strains in iron-metabolism genes. Each panel shows the coding variants private to Malaysian yeast in one iron-metabolism protein, with the variants denoted as stars in a cartoon of the protein’s domain organization at top, and a protein-coding alignment below for the region containing the Malaysian variant, using sequence and strain identifiers from (Liti et al. 2009). (A) Yap5p; (B) Ccc1p; (C) Aft1p. Non-synonymous coding variants private to the Malaysian population are indicated in red. bZIP_1, basic leucine zipper domain, PFAM clan 1; VIT1, vacuolar iron transport domain; AFT, activator of iron transcription domain; NLS, nuclear localization signal.

Supplementary Figure 2. Regulatory impact of variation in AFT1, CCC1, and YAP5 between Malaysian and wine/European yeast. Data are as in Figure 4 of the main text except that distributions of expression measurements across experimental replicates are shown. Each panel reports expression of iron-starvation (Gasch et al. 2004) and iron-resistance (Lin et al. 2011; Pimentel et al. 2012) genes as ratios of the levels measured in two strains derived from a Malaysian (UWOPS03.461.4) and a wine/European strain (BC187). At left in each row, each bar reports the median across replicates (n = 2) of the ratio of expression of the indicated gene between a Malaysian homozygote and a wine/European homozygote. In each remaining panel, each bar reports the median across replicates (CCC1, n = 8; AFT1, n = 4; YAP5, n = 4) of the ratio of expression of the indicated gene between a Malaysian-wine/European hemizygote bearing the Malaysian allele of the indicated variant locus, and the hemizygote bearing the wine/European allele. Error bars report 95% confidence intervals. (A) Iron-starvation genes in synthetic complete medium. (B) Iron-resistance genes in synthetic complete medium supplemented with 5 mM FeSO4. Raw data are reported in Supplementary Data Sets 7 and 8.
Supplementary table captions

Supplementary Table 1. RNA-seq statistics.

Supplementary Table 2. Strains used in this work.
Supplementary data sets

Supplementary Data Set 1. Single-nucleotide polymorphisms in Malaysian and wine/European coding sequences inferred from RNA-seq. Each file reports single-nucleotide polymorphisms (SNPs) in open reading frames of one yeast strain, called with respect to the genome of the reference laboratory strain S288C (www.yeastgenome.org). In a given file, each row reports genotype information at one SNP position: the first column gives the chromosome identifier, the second column gives the position on the indicated chromosome, the third and fourth columns give the base present in the reference genome and the strain of interest, respectively, and the fifth column gives the Phred score of the genotype call.

Supplementary Data Set 2. Expression profiles of Malaysian and wine/European yeast, measured by RNA-seq. Each row reports a comparison of the expression of one gene between homozygous Malaysian (UWOPS03.461.4, UWOPS05.217.3, and UWOPS05.227.2) and wine/European (BC187 and RM11-1) yeast strains grown in rich medium. For a given row, the second through sixth columns report raw counts of RNA-seq reads mapped to the indicated gene in libraries from the indicated strain. The remaining columns report the results of normalization and statistical testing from the DEseq software suite (ANDERS and HUBER 2010): mean, average normalized expression of the indicated gene over all strains; WE_mean, average normalized expression across wine/European strains; M_mean, average normalized expression across Malaysian strains; foldChange and log2FoldChange, ratio and log$_2$ ratio, respectively, of the normalized mean expression in Malaysian strains relative to wine/European strains; pval, p-value assessing significance, in a negative binomial-based test, of the differential expression between populations; padj, corrected p-value by the Benjamini-Hochberg method.

Supplementary Data Set 3. Growth attributes, in synthetic complete medium and in high iron, of Malaysian and wine/European yeast, their hybrids, and reciprocal hemizygotes for AFT1, YAP5, and CCC1. Each row reports growth measurements, fitted to growth curves using the method of (WARRINGER et al. 2011), from one yeast culture in complete synthetic medium mock-treated (CSM) or treated with 5 mM FeSO$_4$. Strain names are as in Supplementary Table 2. Hybrid, wild-type diploid formed by a mating between a Malaysian and a wine/European strain. Each identifier of the form xxx/XXX-Malaysian denotes a reciprocal hemizygote bearing only the Malaysian allele of the gene of interest, and xxx/XXX-wineEuropean denotes a hemizygote bearing only the wine/European allele. Doubling time, the inverse of the slope of a line fitted to the relationship between time (in hours) and cell density during log-phase growth; lag time, the x-intercept of a line fitted to the relationship between time (in hours) and cell density between inoculation and the onset of log-phase growth; efficiency, the ratio between the cell density at the end of a culture and the density at the start.

Supplementary Data Set 4. Growth attributes, in synthetic complete medium and in high iron, of a panel of environmental yeast isolates. Data are as in
Supplementary Data Set 3 except that wild-type isolates from each of the yeast populations defined in (Liti et al. 2009) were analyzed.

**Supplementary Data Set 5.** Expression profiles of hybrid strains formed by a mating between Malaysian and wine/European yeast, measured by RNA-seq. Each row reports comparisons of the expression of the Malaysian allele of a gene to that of the wine/European allele of a gene, in hybrids formed by matings of Malaysian and wine/European strains. For a given gene, the second through sixth columns report raw counts of RNA-seq reads mapped to the indicated allele (M, Malaysian; WE, wine/European) in one hybrid strain, with identifiers as in Supplementary Table 2: YHL058, the Malaysian UWOPS03.461.4 mated to the wine/European RM11; YHL063, the Malaysian UWOPS03.461.4 mated to the wine/European BC187; YHL065, UWOPS05.217.3 mated to BC187. The remaining columns report the results of normalization and statistical testing from the DEseq software suite (Anders and Huber 2010): mean, average normalized expression of the indicated gene over all strains and alleles; WE_mean, average normalized expression across the wine/European alleles of all strains; M_mean, average normalized expression of the Malaysian alleles across all strains; foldChange and log2FoldChange, ratio and log2 ratio of the normalized mean expression of the Malaysian allele across all strains relative to that of the wine/European allele; pval, p-value assessing significance, in a negative binomial-based test, of differential allele-specific expression considering measures across the strains as replicates; padj, corrected p-value by the Benjamini-Hochberg method.

**Supplementary Data Set 6.** Directional cis-regulatory variation between Malaysian and wine/European yeast in co-regulated gene groups. Each row reports the results of a test for directional coherence of cis-regulatory variation between Malaysian and wine/European yeast in one group of functionally related genes, measured using reads uniquely mapped to each parent’s allele of a given gene in turn, in hybrids formed by mating Malaysian and wine/European strains. Group, identifier of regulon from (Gasch et al. 2004) or Gene Ontology term. Upregulated, population with elevated expression; gene groups with an average expression difference of 0 between the populations are denoted with NA. 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 log2 ratio of expression of the Malaysian allele of a given gene to expression of the wine/European allele, assessed using all isolates from each population and corrected for multiple testing with the Benjamini-Hochberg method. An additional test, not shown, for directional cis-regulatory variation across genes in a ccc1 laboratory strain compared to wild-type during high-iron treatment (Lin et al. 2011) yielded a nominal p-value of 0.12.

**Supplementary Data Set 7.** Expression of iron-starvation genes in Malaysian and wine/European parent strains and reciprocal hemizygotes in synthetic complete medium, measured by quantitative PCR. Each column reports comparisons between the effects of Malaysian and wine/European genotypes on expression of an iron-starvation gene during growth in synthetic complete medium. The first six rows report expression in a Malaysian (UWOPS03.461.4) and a wine/European strain.
(BC187). The remaining rows report expression in reciprocal hemizygote pairs in the UWOPS03.461.4 x BC187 background. Each identifier of the form xxx/XXX-Malaysian denotes a reciprocal hemizygote bearing only the Malaysian allele of the gene of interest, and xxx/XXX-wineEuropean denotes a hemizygote bearing only the wine/European allele. In row headings, each numerical value represents one biological replicate, and each row reports one technical replicate.

Supplementary Data Set 8. Expression of iron-resistance genes in Malaysian and wine/European parent strains and reciprocal hemizygotes in high-iron conditions, measured by quantitative PCR. Data are as in Supplementary Data Set 7 except that strains were cultured in synthetic complete medium supplemented with 5 mM FeSO₄, and iron-resistance gene expression was measured.
References


Li, H., B. Handsaker, A. Wyssker, T. Fennell, J. Ruan et al., 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.


Shapiro, M. D., M. E. Marks, C. L. Peichel, B. K. Blackman, K. S. Nereng et al., 2004 Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428: 717-723.


Table 1. Directional expression divergence between Malaysian and wine/European yeast in co-regulated gene groups.

<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>0.0078</td>
</tr>
<tr>
<td>Node 36</td>
<td>Enzymes</td>
<td>Malaysian</td>
<td>0.0156</td>
</tr>
<tr>
<td>GO:0007039</td>
<td>Vacuolar protein catabolic process</td>
<td>Malaysian</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>GO:0006519</td>
<td>Cellular amino acid metabolic process</td>
<td>Wine/European</td>
<td>0.0100</td>
</tr>
<tr>
<td>GO:0006725</td>
<td>Cellular aromatic acid metabolic process</td>
<td>Wine/European</td>
<td>0.0133</td>
</tr>
<tr>
<td>GO:0006811</td>
<td>Ion transport</td>
<td>Wine/European</td>
<td>0.0300</td>
</tr>
<tr>
<td>GO:0006082</td>
<td>Organic acid metabolic process</td>
<td>Wine/European</td>
<td>0.0320</td>
</tr>
<tr>
<td>GO:0006826</td>
<td>Iron ion transport</td>
<td>Wine/European</td>
<td>0.0400</td>
</tr>
<tr>
<td>GO:0055085</td>
<td>Transmembrane transport</td>
<td>Wine/European</td>
<td>0.0400</td>
</tr>
</tbody>
</table>
Table 2. Elevated protein evolutionary rates in iron-toxicity response genes in Malaysian yeast.

<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>0.1469</td>
<td>0.1240</td>
<td>0.2871</td>
</tr>
<tr>
<td>Malaysian</td>
<td>0.2629</td>
<td>0.1650</td>
<td>0.0416</td>
</tr>
<tr>
<td>Sake</td>
<td>0.0594</td>
<td>0.1014</td>
<td>0.7481</td>
</tr>
<tr>
<td>North American</td>
<td>0.0487</td>
<td>0.0822</td>
<td>0.7511</td>
</tr>
<tr>
<td>West African</td>
<td>0.1878</td>
<td>0.1777</td>
<td>0.3861</td>
</tr>
</tbody>
</table>