

Cascading transcriptional effects of a naturally occurring frameshift mutation in *Saccharomyces cerevisiae*

KYLE M. BROWN,* CHRISTIAN R. LANDRY,*† DANIEL L. HARTL* and DUCCIO CAVALIERI‡

*Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

†Current Address: Département de Biochimie, Faculté de Médecine, Université de Montreal, C.P. 6128, Succ. Centre-Ville, Montreal, Quebec, Canada H3C 3J7, ‡Dipartimento di Farmacologia, Università di Firenze Viale Pieraccini 6, 50139 Firenze, Italy

Abstract

Gene-expression variation in natural populations is widespread, and its phenotypic effects can be acted upon by natural selection. Only a few naturally segregating genetic differences associated with expression variation have been identified at the molecular level. We have identified a single nucleotide insertion in a vineyard isolate of *Saccharomyces cerevisiae* that has cascading effects through the gene-expression network. This allele is responsible for about 45% (103/230) of the genes that show differential gene expression among the homozygous diploid progeny produced by a vineyard isolate. Using isogenic laboratory strains, we confirm that this allele causes dramatic differences in gene-expression levels of key genes involved in amino acid biosynthesis. The mutation is a frameshift mutation in a mononucleotide run of eight consecutive T's in the coding region of the gene *SSY1*, which encodes a key component of a plasma-membrane sensor of extracellular amino acids. The potentially high rate of replication slippage of this mononucleotide repeat, combined with its relatively mild effects on growth rate in heterozygous genotypes, is sufficient to account for the persistence of this phenotype at low frequencies in natural populations.

Keywords: ecological genomics, fitness, gene expression, gene networks, population genetics, wine yeast

Received 12 January 2008; revision received 11 March 2008; accepted 13 March 2008

Introduction

Differences in gene expression are important sources of phenotypic variation. Early studies demonstrated that genome-wide variation in gene expression is abundant across taxonomic groups, including between species (Hsieh *et al.* 2003; Ranz *et al.* 2003; Rifkin *et al.* 2003), within populations (Jin *et al.* 2001; Townsend *et al.* 2003; Aubin-Horth *et al.* 2005; Whitehead & Crawford 2006; Clark & Townsend 2007) and between progeny (Cavalieri *et al.* 2000). Comparisons of expression data have allowed inferences about the underlying evolutionary forces that shape expression variation on different evolutionary time scales (e.g. Jordan *et al.* 2005; Lemos *et al.* 2005, 2007). While a few studies of gene-expression divergence between species and polymorphism within species suggest a role for directional selection on

transcript abundance (Rifkin *et al.* 2003; Landry *et al.* 2007a), the emerging consensus is that stabilizing selection limits the amount of gene-expression polymorphism in nature by selecting for phenotypes with intermediate levels of expression (Lemos *et al.* 2005; Denver *et al.* 2005).

The functional consequences of genome-wide expression variation have mainly been addressed indirectly. Such studies have compared how the inferred relative strengths of the selective forces map differentially among groups of functionally related genes as defined, for example, by gene ontology categories (Lemos *et al.* 2005). Furthermore, the identification of the evolutionary forces acting on gene-expression polymorphism has, so far, mostly relied on the use of the theory developed for the evolution of quantitative characters. However, the genetic bases of gene-expression phenotypes can be simple in some cases and very complex in others (Gibson *et al.* 2004; Brem & Kruglyak 2005), thus violating some of the assumptions of classical quantitative genetics. A more direct approach would be first to identify the molecular basis of a gene-expression phenotype and

Data deposition: GSM248644–GSM248649.

Correspondence: Kyle M. Brown, Fax: +1 617 496 5854; E-mail: kmbrown@fas.harvard.edu

then to investigate the forces acting on this variation. However, the molecular underpinnings of phenotypic variation, including gene-expression differences, are often difficult to identify. Even as techniques such as genetic mapping of quantitative trait loci (QTL) reveal regions of the genome that contribute significantly to phenotypic variation (Bradshaw *et al.* 1995), it is often difficult to narrow the region of genetic causation to a single gene or nucleotide (Erickson *et al.* 2004; Ron & Weller 2007; but see: Bradshaw & Schemske 2003; Ben-Ari *et al.* 2006).

The budding yeast *Saccharomyces cerevisiae* has been at the forefront of the methods, technologies and statistical approaches used to dissect the transcriptome genetically (Brem *et al.* 2002; Steinmetz *et al.* 2002). This unicellular fungus has also pioneered the field of ecological genomics, whereby variation in gene expression among natural isolates is mapped onto known and well defined cellular, metabolic and regulatory networks (Cavalieri *et al.* 2000; Fay *et al.* 2004; Landry *et al.* 2006a). The wealth of functional genomics and molecular genetics resources available for this model species rank it highly among the model organisms best suited to tackle the question of how gene-expression phenotypes may affect organismal fitness.

Using the functional genomic tools then available in yeast, Cavalieri *et al.* (2000) discovered progeny from a single vineyard strain of *S. cerevisiae* that displayed widely divergent gene-expression phenotypes. Originally identified because of their distinct colony morphologies (Cavalieri *et al.* 2000), diploid progeny from a homothallic vineyard isolate designated 'M28' differed in gene-expression levels at 230 loci, or 3.6% of all genes, based on a Bayesian analysis (Landry *et al.* 2006b; see Materials and Methods for analysis of statistical power). While these expression differences did not cosegregate with colony morphology (KMB and DC, unpublished results), the expression phenotypes mapped to specific metabolic pathways, and a large fraction appeared to segregate together as a single Mendelian trait, suggesting a simple genetic basis (Cavalieri *et al.* 2000; Landry *et al.* 2006b).

We set out to identify the genetic basis of this expression pattern and to evaluate the effects of the phenotype on growth rate in laboratory culture. We found that a single nucleotide insertion in *SSY1*, a gene encoding a component of a plasma-membrane sensor of extracellular amino acids, is responsible for 45% of the gene-expression differences among the progeny of M28, the parental strain studied by Cavalieri *et al.* (2000). By introducing this mutation into isogenic strains of *S. cerevisiae*, we showed that this change alone is sufficient to reproduce the expression phenotype affecting amino acid biosynthesis and transport that segregates in M28. We compared growth rates in laboratory culture among the progeny of the vineyard strain M28 and of isogenic strains containing the mutation. We observed that detrimental effects on growth rate in the heterozygous

genotype are undetectable under our conditions, although they are quite large in the homozygous mutant. Given that this allele results from the expansion of an unstable mononucleotide repeat and has also been found in another independent isolate of *S. cerevisiae*, we hypothesize that this allele may persist at low frequencies in wild populations of *S. cerevisiae* owing to its high mutation rate balanced against its modest selective effects when heterozygous.

Materials and methods

Saccharomyces cerevisiae strains

Saccharomyces cerevisiae strains M28 and M57 were isolated in Montalcino (Tuscany), Italy, from damaged grapes (Cavalieri *et al.* 1998), a typical source of vineyard yeasts (Mortimer & Polsinelli 1999). As described in Cavalieri *et al.* (2000), strains 4A and 4D (originally referred to as S1 and F₂) are derived from individual meiotic segregants from a single tetrad. Because M28 resembles about 70% of vineyard isolates in being homothallic, all progeny from sporulation are homozygous diploids (Mortimer 2000).

Because natural isolates have no auxotrophic markers for use in genetic manipulation, we created a *ura3* derivative of 4D. After sporulating 4D on standard sporulation medium, we plated spores on 4,4,4 fluororotic acid (FOA), a technique that selects for *ura3* mutants (loss of function). The presence of *ura3* was verified by confirming that strains could not grow on medium lacking uracil. Furthermore, we showed that putative *ura3* 4D derivatives could grow in the absence of uracil after being transformed with pRS 416, a plasmid carrying *URA3* (ATCC No. 87521; Sikorski & Hieter 1989). S288C derivatives FY2, FY3 and FY4 are described in Winston *et al.* (1995).

Complementation tests

We obtained plasmids with *URA3* markers carrying the genes encoding *SSY1*, *PTR3* and *SSY5* from Peter O. Ljungdahl (Forsberg & Ljungdahl 2001). We used these plasmids to transform *ura3* 4D using medium without uracil to select for transformants. Transformations were performed according to standard methods (Adams *et al.* 1997). To determine whether the wild-type gene introduced via the plasmid complemented the mutant phenotype, we tested the resistance of the transformants to 5,5,5-tri-fluoro-DL-leucine (TFL; Sigma-Aldrich) at 50 µg/mL on plates.

Sequencing

The *SSY1* locus and flanking regions in the 4D strain (4D; referred to as F₂ in Cavalieri *et al.* 2000) were sequenced using primers designed from the published S288C sequence (*Saccharomyces* GENOME Database; www.yeastgenome.org).

The entire locus and flanking regions including flanking regions and surrounding genes (*SAC3* and *YDR161w*) were sequenced in order to identify the mutation (**Region 1:** TTACTCCTGTGCTAGATGG = Forward, ATTCTTAGCAAGGCTGGACG = Reverse; **Region 2:** CGTCCAGCCTTGCTAAGAAT = Forward, CGTCCAGCCTTGCTAAGAAT = Reverse; **Region 3:** TGCCCCTCGATAGATACTCC = Forward, CCCTGCTGTTACTCCCTTTG = Reverse; **Region 4:** TGTATGCTTGCTCTTCCTG = Forward, TAGTAGTCTTGGATCCAC = Reverse; **Region 5:** CCATTGTAAGTATGATATC = Forward, TAGATATGGTTGGAATGGCG = Reverse; **Region 6:** TTTGCGATTCTATTACGCC = Forward, CCTACTAAGGTCATCTCATCG = Reverse; **Region 7:** GTTACCTGGCTGATTTAGGG = Forward, TCAGTATTACATTGTTTGGC = Reverse). Sequencing reactions were performed using Big Dye version 3 in a 3100 Genetic Analyser (Applied Biosystems). Multiple sequence alignments comparing 4D sequences with published S288C sequences were created using SEQUENCHER (version 4.2.2, GENE Codes Corporation). The frameshift mutation responsible for the *ssy1^{T9}* allele was found in Region 6.

To ensure that the identified mutation was transcribed into RNA, we sequenced the *SSY1* mRNA from both 4D and S288C. Total RNA was extracted using the methods described below. Gene-specific primers for the mutated region were used to create *SSY1* complementary DNA (cDNA) from total RNA according to protocols described for Superscript II Reverse Transcriptase (Invitrogen). Polymerase chain reaction (PCR) products for the mutated region were created and sequenced according to the procedure described above. Additional strains from the same M28 tetrad (see Cavalieri *et al.* 2000) and another filigreed strain 'M57' were also sequenced at *SSY1* according to the above methods.

Transformation of mutant *SSY1* into S288C using TFL

As the complementation tests described above showed that the *ssy1^{T9}* allele of 4D conferred resistance to TFL, we were able to directly transform a laboratory strain with this allele. Briefly, PCR products were created from the mutated region of *SSY1* in 4D (Forward = CCCAAGTTTCCTCAA GTACATTCTATCTCAGC, Reverse = TTTTGTATTAACCTCTCACACTGCTCTTCTGC). These PCR products were used to transform the S288C strain, FY2 (Winston *et al.* 1995; Brachmann *et al.* 1998) via homologous recombination using standard yeast PCR product transformation procedures (Adams *et al.* 1997). Transformants were selected on TFL (50 µg/mL), uracil (URA) + plates and the appropriate homologous recombination was confirmed by sequencing.

Knockout of *SSY1* in S288C using *URA3*

In order to compare the 4D mutation with a gene knockout, we constructed an *SSY1* knockout in an S288C strain, FY2.

While this knockout is lethal in a *leu* background (Klasson *et al.* 1999; *Saccharomyces* GENOME Database), knocking out this gene in *LEU* S288C results in a viable strain. Approximately 80 base-pair (bp) primers were designed with the 3'-most 20 bp homologous to flanking regions around the *URA3* locus in S288C (Forward = GATAATCACC CATTCTTATTACTGTCGGCGTGCCATCCTTTTCTCTAAAG CTCGGCGGAACATAACATCACACTTGCTGG, Reverse = CG GTCCAGTCAGTTCTGTCCATTTCCCTCCTCCCACTGTCC ATATCTAACTGATCTAGTCTGTTCCGTTTACTTGTCCG). The 5'-most 60 bp of each primer were designed to match flanking regions around *SSY1*. PCR products of *URA3* were created using these primers and then transformed into FY2. Knockouts of *SSY1* were selected on media lacking uracil. Additional PCR primers were designed to amplify from the DNA of putative transformants only if the knockout had been successful (Forward = TGTTACTCCT GTGCTAGATGGG, Reverse = TGCTCTAGCGTTACCA CTGC). Resulting PCR products were sequenced to confirm the replacement of *SSY1* with *URA3*.

DNA microarrays

Cells were grown in 5 mL of YPD (1% yeast extract, 2% peptone, 2% dextrose) or MM (0.67% yeast nitrogen base without amino acids, 2% dextrose) overnight and diluted to an OD₆₀₀ of 0.1 in 40 mL of YPD in the morning and left to grow to an OD₆₀₀ of ~0.8 at 30 °C and 225 r.p.m. Cultures were then centrifuged at 3500 r.p.m. for 20 min at room temperature and the pellets were flash frozen in liquid nitrogen. Frozen cell pellets were resuspended and extracted with hot acidic phenol/chloroform extraction. Total RNA was ethanol precipitated, washed and resuspended in TE buffer. RNA quality was confirmed by spectrophotometric analyses with A₂₆₀/A₂₈₀ ratios ~2. cDNA synthesis, labelling and hybridization were performed using 3DNA ARRAY 50 Kit version 2 (Genisphere Inc) according to manufacturer protocol. The samples were hybridized on arrays containing 6388 probes (Qiagen Operon) printed on poly L-lysine coated slides (Erie) according to standard protocols (www.microarray.org).

The arrays were scanned on an Axon GENEPLEX 4000B Scanner (Axons Instrument, Molecular Devices) and the images were analyzed using GENEPLEX Pro 5 (Axons Instrument, Molecular Devices). Spots of bad quality were flagged manually and eliminated from the downstream analyses. We also discarded spots that had a foreground intensity smaller than the background intensity plus two standard deviations for both channels. The raw intensities were normalized using the print-tip loess method implemented in the Limma package of the R statistics software (Smyth & Speed 2003). This method ensures that the ratios of expression are independent of the mean intensities. Statistical analysis of differential gene expression was the

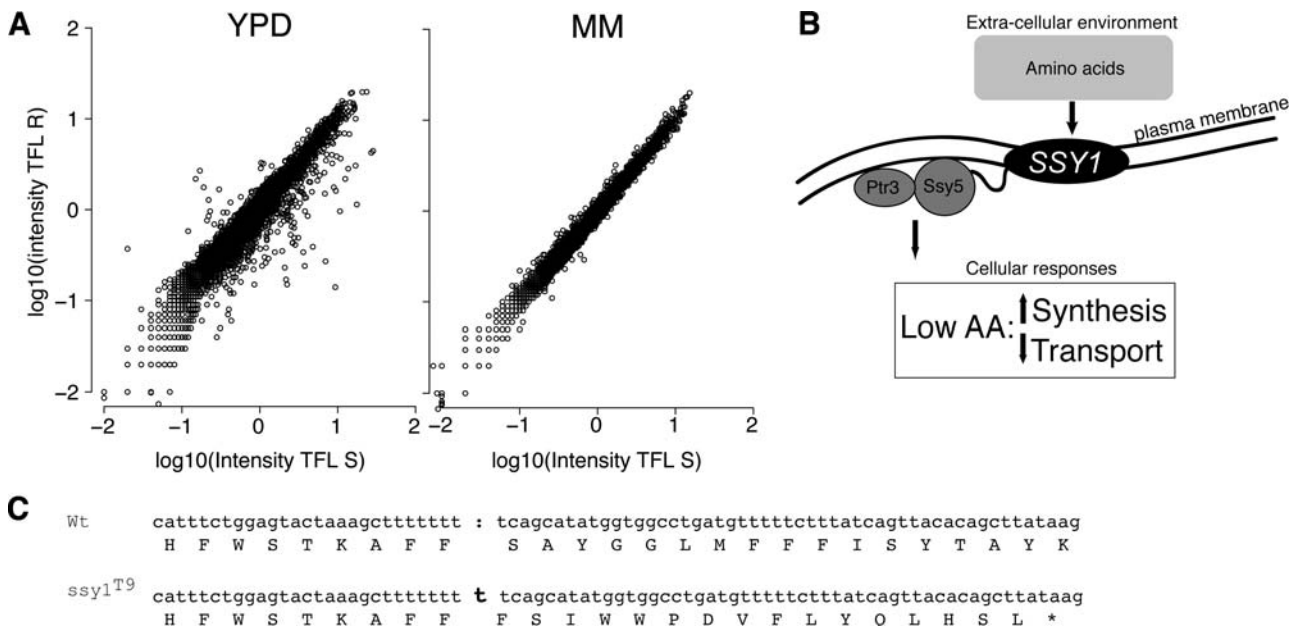


Fig. 1 (A) A comparison of gene expression patterns in a TFL-resistant M28 segregant (4D) and a TFL-sensitive M28 segregant (4A) revealing more differentially expressed genes between TFL-resistant and TFL-sensitive in YPD medium than in minimal medium (MM). Points represent average normalized spot intensities across two replicate microarrays for the resistant strain vs. the sensitive strain. (B) A representation of the SPS complex and the role of Ssy1p. (C) Comparison of *SSY1* sequences from wild-type and *ssy1^{T9}* strains. Sequences shown represent nucleotides 2332–2410 and the corresponding predicted amino acid sequences, amino acids 778–803. Because of the frameshift mutation in *ssy1^{T9}*, the mutant protein is predicted to terminate after translating only 802 amino acids (2409 nucleotides), 50 amino acids fewer than the wild type.

performed using BAGEL (version 3.62 for Windows) with the default parameters (Townsend & Hartl 2002; see Table S2, Supplementary material for results). TFL resistant strains vs. sensitive strains gene-expression comparisons, described in Fig. 1A, were performed as described in Cavalieri *et al.* (2000). While the specific number of gene-expression differences depends on the statistical power of the microarray methodology, arrays whose data was presented in Fig. 1A were performed to test the effect of amino acids in the growth media on genome-wide expression patterns between TFL-sensitive and -resistant M28 progeny. Therefore, they were not analyzed to detect statistically significant expression differences for individual genes but to confirm that the presence of amino acids affected the gene expression differences.

For all other array data, we tested the power of our methods to detect gene-expression differences by determining the gene-expression level difference (fold-change) at which there is a 50% chance of a significance call (Townsend 2004; Clark & Townsend 2007). This statistic is known as the GEL50. Using our stringent criteria of non-overlapping 95% Bayesian credible intervals, results from arrays detailed in Landry *et al.* (2006b) have higher GEL50 values than previously reported (Clark & Townsend 2007). Arrays comparing the TFL sensitive segregant M28 4A

with TFL resistant segregants M28 4C and M28 4D had GEL50 values of 3.44 and 3.58, respectively. Arrays comparing TFL sensitive M28 4B with 4C and 4D had somewhat higher power to detect differences (2.56 and 2.85, respectively). In contrast, arrays comparing *SSY1* and *ssy1^{T9}* strains of S288C had a GEL50 value of 1.24. The higher statistical power gives these comparisons a greater probability of detecting gene expression differences caused by the presence of the *ssy1^{T9}* allele.

Construction of S288c diploids

In order to simulate the effects that the *ssy1^{T9}* allele might have in isolation on wild growing strains, we mated haploid S288C strains to create diploid strains containing *SSY1* mutations and knockouts in heterozygous and homozygous forms. Two selectable markers, *kanMX6* and *URA3*, ensured that only mated cells would survive on plates containing G418 and lacking uracil. (Strains containing *kanMX6* are resistant to G418, which we obtained from Sigma-Aldrich.) Specifically, two strains were constructed from FY3 (Winston *et al.* 1995). In one transformation, we replaced the *HO* locus with the *kanMX6* module from the pFA6a-*kanMX6* PLASMID (Longtine *et al.* 1998; see above for PCR transformation protocol, Forward = CTATTACAACACTATTAG

CTCTAAATCCATATCCTCATAAGCAGCAATCAATTCT
ATCTATACCCGGATCCCCGGGTTAATTAA, Reverse = TCAT
CCAAAATATTAATTTTACTTTTATTACATACAACTTTT
TAAACTAATATACACATTGAATTCGAGCTCGTTAAAC).
Another FY3 derivative was constructed by knocking out
SSY1 and replacing it with the same cassette (Forward =
GATAATCACCAATTCCTTATTACTGTCCGGCGTGCCATCC
TTTTCTCTAAAGCTCGGCGGACGGATCCCCGGGTTAATT
AA, Reverse = CCGTCCAGTCAGTTCTGTCCATTTCCCT
CCTCCCACTGTCCATATCTAACTGATCTAGTCAATT
CGAGCTCGTTAAAC). The FY2 *ssy1^{T9}* strain, described
above, was transformed to replace its HO locus with
kanMX6. Four crosses were performed to create the necessary
genotypes for growth rate experiments (Cross A = FY3
ho::kanMX6 × FY5, Cross B = FY3 *ho::kanMX6* × FY2 *ssy1::URA3*,
Cross C = FY3 *ssy1::kanMX6* × FY2 *ssy1::URA3*, Cross D = FY2
ssy1^{T9} × FY4). All resulting diploids were heterozygous
for both selectable markers (*URA3* and *kanMX6*). Cross A
resulted in a homozygous wild-type *SSY1* strain (denoted
wt/wt in Fig. 4). Cross B and Cross D resulted in hetero-
zygote strains for the knockout and 4D *ssy1* allele,
respectively (denoted wt/KO and wt/^{T9}). Cross C resulted
in a homozygous *ssy1* knockout strain (denoted KO/KO).

Diploid strain maximum growth rate assays

The maximum growth rates of S288C derivative diploids
and M28 derivative strains were assayed using optical
density measurements obtained from a Bioscreen C micro-
biological workstation (Thermo Labsystems). Following
Joseph & Hall (2004), single colonies were picked from
plates and grown overnight in liquid YPD or MM with
uracil (20 mg/L). Optical densities of each overnight culture
were recorded and used to dilute cultures to approximately
10⁵ cells/mL. 150 µL of these dilutions were loaded into
each well of the Bioscreen C plate. Two plates of one hundred
wells, containing at least four replicates of each strain in
each media type, were assayed for their optical density every
15 min for 2 days. All wells reached stationary phase, as
determined by optical density plots, after 18 h.

Using code written in *R* (version 2.2.1, The R Foundation
for Statistical Computing), we calculated a least-squares
regression for log absorbance vs. time for a 3.25-h sliding
window over the length of the growth curve. The maximum
growth rate from each well was designated to be the
coefficient from the regression with the largest slope.

Statistical analyses were performed in *R* and *SPSS* (*SPSS*
for Windows, version 13.0, SPSS Inc.). A nested ANOVA com-
paring growth rates of replicates in different plates revealed
no significant plate effect. Replicates of similar strains and
media across the two plates were pooled for subsequent
statistical analyses. Pair-wise *t*-tests were used to compare
the maximum growth rates of different strains in the same
media type.

T-test power simulations

We verified that samples of replicate growth-rate experiments
performed in YPD had smaller variances than those in MM
by using Bartlett's test for homogeneity of variances per-
formed in *R*. We subtracted each replicate growth rate
value from its corresponding sample mean to remove any
strain effect but preserve the variances of each sample.
Strains grown in YPD had statistically indistinguishable
variances (Bartlett's *K*-squared = 0.7708, *P* = 0.9928), as did
those grown in MM (Bartlett's *K*-squared = 0.6287, d.f. = 6,
P = 0.996). We then pooled replicates grown in YPD and
compared their variance with those grown in MM. Growth-
rate experiments performed in YPD had significantly lower
variances than those grown in MM (Bartlett's *K*-squared
= 92.0377, d.f. = 1, *P* < 2.2 × 10⁻¹⁶).

We performed two sets of simulations to test the power
of *t*-tests to determine a difference in parametric mean
growth rate between two strains. Using the average standard
deviation for strains growing in YPD, we sampled from two
normal distributions whose means differed by a predeter-
mined amount to create two samples of eight observations.
We then tested whether a *t*-test would reject the null
hypothesis at the significance level of 0.05. We ran simula-
tions that tested parametric differences in growth rate
over two orders of magnitude (between 0.38% and 38%
difference between means compared). One thousand itera-
tions of the simulation were performed for each difference
in mean. Identical simulations were performed using the
average standard deviation for strains growing in MM (See
Fig. S2, Supplementary material, for simulation results).

Results

SSY1 is the cause of gene expression variation

We identified *SSY1* as the source of gene-expression variation
among the progeny of strain M28 by identifying the genetic
basis of resistance to TFL, a toxic leucine analogue. Like the
gene-expression differences, TFL resistance segregates 2 :
2 among progeny derived from single tetrads of M28.
Previous work had demonstrated that TFL-resistant progeny
of M28 differentially expressed over 200 genes in com-
parison to either the M28 parent or TFL-sensitive progeny
(Cavalieri *et al.* 2000; Townsend *et al.* 2003; Landry *et al.*
2006b). These expression experiments therefore indicated
that the TFL-resistance phenotype cosegregated with the
gene-expression phenotype. In comparison with the parent
or TFL-sensitive progeny, TFL-resistant M28 progeny up-
regulated many genes for amino acid biosynthesis while
they down-regulated amino acid transporter and permease
genes (Cavalieri *et al.* 2000; Townsend *et al.* 2003; Landry
et al. 2006b). This suggested that a defect in amino acid
synthesis or utilization was responsible for the expression

phenotype. Since TFL must be transported into the cells via amino acid permeases and transporters, TFL resistance suggested a defect in amino acid transports across the plasma membrane. In fact, differences in gene expression between TFL-resistant and TFL-sensitive M28 segregants were similar to differences between strains growing in the presence and absence of amino acids (Wodicka *et al.* 1997; Landry *et al.* 2006b).

Based on this information, we tested the hypothesis that the presence of amino acids affected the expression differences. Comparison of the expression profile of a representative pair of TFL-resistant and -sensitive progeny from a single tetrad from M28 in the presence and absence of amino acids confirmed this prediction (Fig. 1A). A preliminary set of microarrays suggested that more genes exhibited large expression differences in YPD medium, which is rich in amino acids, than in minimal medium (MM), which lacks amino acids. For example, 286 genes are differentially expressed more than two-fold between the TFL-resistant segregant 4D compared to its TFL-sensitive cosegregant 4B when both segregants are grown in YPD (Fig. 1A). However, growing these same strains in MM eliminated all but 12 of these two-fold expression differences (Fig. 1A, see Materials and Methods). Many of the expression differences between the TFL-resistant and TFL-sensitive segregants seemed to be related to the presence of amino acids in the external environment and confirm that the expression differences among these segregants are likely related to the use or sensing of amino acids.

Using data from these expression experiments along with the knowledge that TFL needs to be imported into the cells to be toxic, we hypothesized that the expression differences may result from defects in the sensing of extracellular amino acids. One particularly well-studied complex of amino-acid-sensing genes is the SPS nutrient sensor, which includes the products of the genes *SSY1*, *PTR3*, and *SSY5* (Fig. 1B, Poulsen *et al.* 2005). In the presence of amino acids in the environment, the SPS complex triggers downstream pathways that lead to the expression of amino acid transporters and permeases with concomitant down-regulation of amino acid biosynthetic pathways. SPS therefore plays a fundamental role at the interface between the cells and the environment. The role of this nutrient sensor has been extensively studied, and the expression in the 4D segregant might be expected from a strain with a mutation in one of these loci (Klasson *et al.* 1999; Forsberg & Ljungdahl 2001; Forsberg *et al.* 2001; Eckert-Boulet *et al.* 2004).

To ascertain whether a mutation in a component of the SPS extracellular amino acid sensory system might be responsible for the metabolic phenotype, we tested if the wild-type alleles of SPS genes could restore the TFL sensitivity of the 4D segregant by introducing the wild-type alleles of these genes into the 4D background (complementation tests). We observed that only the *SSY1*-containing plasmid

restored TFL sensitivity to 4D (see Materials and Methods). While not conclusive, this observation was consistent with the hypothesis that a mutation effecting TFL resistance, and potentially the gene-expression phenotype, may be contained within *SSY1*.

We therefore sequenced the complete coding and flanking region of *SSY1* in 4D and identified a mutation in the coding region. The mutation identified in 4D is a frameshift mutation in the distal part of the coding region. We refer to the *ssy1* allele in the TFL-resistant segregant 4D as *ssy1^{T9}* because the insertion adds a ninth thymidylate in a mononucleotide repeat normally containing eight (Fig. 1C). The protein encoded by *ssy1^{T9}* is predicted to be 50 amino acids shorter than the wild type, with 15 of the 16 amino acids encoded after the frameshift in *ssy1^{T9}* predicted to be different from those at corresponding sites in the wild-type polypeptide. We verified that *ssy1^{T9}* allele was transcribed by sequencing mRNA transcripts from 4D. We confirmed this result by sequencing the three independent segregants isolated from other tetrads of M28. As predicted by the TFL-resistance data reported by Cavalieri *et al.* (2000), the progeny that demonstrated TFL resistance had the *ssy1^{T9}* allele while those that were TFL sensitive did not.

Additionally, we observed this mutation in the genetically distinct vineyard isolate M57. Previously, sporulation of M57 had been shown to yield 2 : 2 segregation of TFL resistance: TFL sensitivity (DC, unpublished results). Two TFL-resistant segregants from M57 had *ssy1^{T9}* sequences identical to that in 4D. While both strains were isolated from Tuscany, Aa *et al.* (2006) sequenced four genes in M28 and M57 and demonstrated divergence in gene sequence at three of these loci. Additionally, results of gene expression profiling from Townsend *et al.* (2003) showed that 203 genes are differentially expressed to a significant degree between M28 and M57, suggesting that these strains are not more closely related than other pairs of strains from the same population. M57 therefore represents an independent sample and suggests that *ssy1^{T9}* segregates in natural populations.

Frameshift allele, ssy1^{T9}, is sufficient to create gene-expression phenotype

To test whether the *ssy1^{T9}* allele is the major genetic determinant of the gene-expression phenotype of the M28 progeny, we created isogenic S288c-derived strains with either the *ssy1^{T9}* or wild-type *SSY1* allele. Because the *ssy1^{T9}* allele confers resistance to the drug TFL (see Materials and Methods), we used full-length PCR products of the *ssy1^{T9}* allele from 4D to transform S288C directly with no additional genetic markers. Microarray analysis from total RNA revealed a similar pattern of expression differences between isogenic strains of S288C with *SSY1* and *ssy1^{T9}* as were present between the TFL-resistant and TFL-sensitive segregants

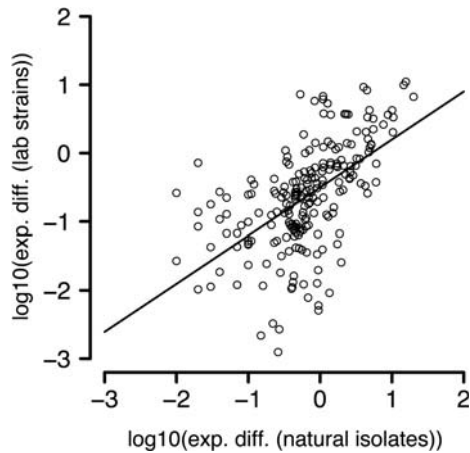


Fig. 2 (A) Relative expression levels of nine genes in the methionine biosynthesis pathway in natural isolates and laboratory strains. Error bars represent the 95% credible intervals. Note that the scales of the relative expression levels are different in each graph. While expression patterns are similar between the two comparisons, differences in relative expression between comparisons of natural and laboratory strains are likely due to differences in microarray technologies and protocols between the two studies. (B) Relative expression levels of five amino acid permeases in natural isolates and laboratory strains. Error bars represent the 95% credible intervals.

from M28. Of the 230 genes differentially expressed between M28 progeny based on nonoverlapping Bayesian credible intervals (Landry *et al.* 2006b), 103 have nonoverlapping credible intervals in comparisons between S288C *SSY1* and S288C *ssy1^{T9}*. We interpret this result to indicate that approximately 45% of the gene-expression differences between 4D and 4A result from this single nucleotide difference. Our arrays comparing S288C strains had higher statistical power than those comparing M28 progeny, indicating that gene-expression differences between the progeny of M28 not represented in comparisons with the S288C strain are likely due to other segregating genetic factors (see Materials and Methods). Further, we assessed the genome-wide expression correlation between 4D and S288C *ssy1^{T9}* using the 230 genes identified as differentially expressed between M28 itself and its 4D segregant. Remarkably, we find that the expression differences in the two comparisons are significantly correlated despite their differing genetic backgrounds (Fig. 2; Pearson correlation coefficient: 0.56, $P < 2.2 \times 10^{-16}$). While the effect of the *ssy1^{T9}* mutation in gene M28 progeny may be different than in S288C laboratory strains, the correlated expression levels presented above suggests that the average effect of this mutation in laboratory and wild strains is similar. Not only are many of the same genes differentially expressed in the isogenic strains as the M28 progeny, but these genes also exhibit similar levels of differential expression in the two comparisons.

Because of the increased statistical power of our lab strains array methodology (see Materials and Methods), we observed 1143 differentially expressed genes between *SSY1* and *ssy1^{T9}* S288C strains. However, many of these genes exhibited much smaller levels of differential expression. (Average fold change 1040 genes differentially expressed between *SSY1* and *ssy1^{T9}* S288C but not in M28 progeny: 1.41. Compare with average fold change in lab strains for the 103 genes differentially expressed in both M28 progeny and S288C strains: 2.31. Wilcoxon rank sum test: $P < 2.2 \times 10^{-16}$). Of these 1143 genes, 853 were included on the original array comparing M28 progeny. Among these 853 genes, those that were not differentially expressed in M28 progeny also have a small level of putative differential expression. (Average fold change among non-differentially expressed genes: 1.35. Compare with average fold change in M28 progeny for 103 genes differentially expressed in natural and lab strain comparisons: 3.26. Wilcoxon rank sum test: $P < 2.2 \times 10^{-16}$).

Even when arrays comparing natural isolate progeny failed to detect a significant difference in expression level, the putative differences in expression level between M28 progeny were still correlated with those of the lab strains. Like the 103 genes with nonoverlapping confidence intervals in both comparisons, the 853 genes that were differentially expressed in S288C strains and present on arrays comparing M28 progeny showed a correlation in their expression differences in these comparisons (Pearson correlation coefficient: 0.40, $P < 2.2 \times 10^{-16}$). In fact, this correlation holds even when the 103 genes differentially expressed in both comparisons are removed (Pearson correlation coefficient: 0.20, $P < 5.03 \times 10^{-8}$).

The expression phenotypes that show a 2 : 2 segregation among the progeny of M28 often map to pathways involved in the biosynthesis of amino acids. The comparison of S288C *SSY1* and S288C *ssy1^{T9}* revealed the similar expression differences in those pathways. For instance, Fig. 3A shows the relative expression levels of genes in the methionine pathway. Other genes involved in amino acid biosynthesis including those in the histidine, valine and leucine pathways also exhibit similar patterns of gene expression differences in the two *ssy1^{T9}* strains (see Table S2, Supplementary material). Further, both 4D and S288C *ssy1^{T9}* exhibit similar down-regulation of key amino-acid transporters and permeases (Fig. 3b). The overexpression we observed in *GAP1* contrasted strikingly to the general pattern of down-regulated transporters in *ssy1^{T9}* (Fig. 3B). However, up-regulated *GAP1* in *SSY1* mutants is expected in light of previous research indicating that a functional *SSY1* locus is necessary to repress this transporter (Klasson *et al.* 1999; Forsberg *et al.* 2001). In total, these data indicate that the *ssy1^{T9}* allele is sufficient to cause virtually all of the gene-expression differences associated with amino acid transport and biosynthesis between the progeny of M28.

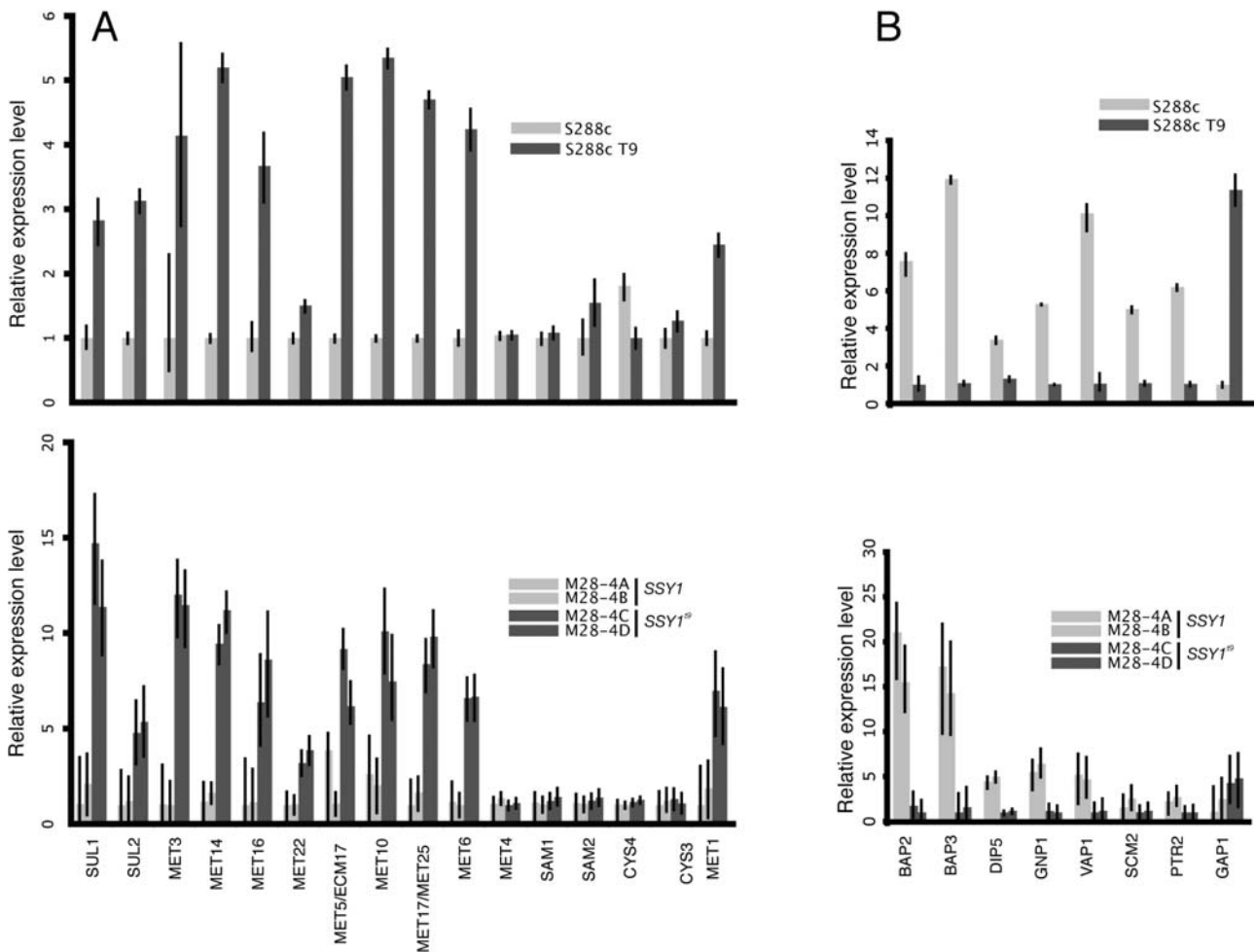


Fig. 3 Correlation of expression differences in genes originally found to be overexpressed between natural isolates (M28 and 4D) with expression differences observed in the laboratory strains (S288C *SSY1* and S288C *ssylT9*). Genes used above were originally found to be differentially expressed in M28 progeny by virtue of having nonoverlapping confidence intervals. Expression differences represent the absolute value of the difference in relative expression between the two compared strains for a given gene. The expression differences between the natural isolates describe a large amount of the variation in expression difference among genes between laboratory strains and the correlation is highly significant (Pearson correlation coefficient: 0.5640047, $P < 2.2 \times 10^{-16}$).

Because the N-terminal domain of Ssy1p is essential for signal transduction to sense extracellular amino acids (Klasson *et al.* 1999), and expression of a truncated polypeptide containing only the N-terminal 276 amino acid residues allows some growth on leucine as sole nitrogen source (Bernard & Andre 2001), we hypothesized that the 800 amino acid polypeptide encoded by the *ssylT9* allele in strains M28 and M57 may retain a fraction of its functions. To test this hypothesis, we extended our gene-expression analysis by comparing the expression pattern of an S288c *ssyl* knockout, referred to as *ssylKO*. We found that only 17 genes are differentially expressed between *ssylKO* and *ssylT9* at a Bayesian posterior probability threshold of 0.9975, which corresponds to a false discovery rate of ~5% (estimated by randomly permuting the data matrix). Most of

these genes are involved in uracil metabolism, the up-regulation of which is likely the result of using *URA3* as a selectable marker. In summary, our results show that *ssylT9* is responsible for the expression phenotype in M28 while having no detectable residual function.

Growth-rate effects of ssylT9 are recessive and environment-dependant

After finding two genetically distinct vineyard isolates heterozygous for *ssylT9* (M28 and M57), we examined the effects of this allele on growth rate in laboratory cultures either when heterozygous or homozygous. We created diploid S288c strains in order to mimic the genetic conditions under which we originally observed this mutation, and introduced

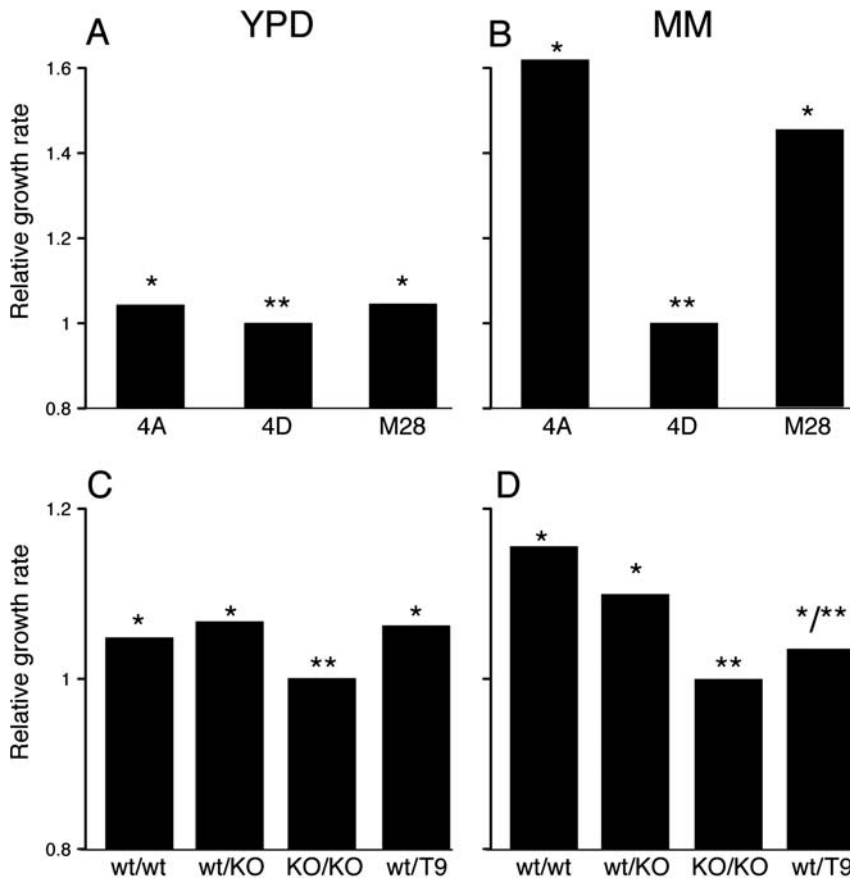


Fig. 4 Relative growth rates of natural isolate (A and B) and diploid laboratory (C and D) strains in media with (YPD, A and C) and without (MM, B and D) amino acids. Growth rates were compared with other closely related strains in the same medium only (e.g. lab strains in MM) and normalized to the slowest growth rate among those three or four strains. Within each part of the figure, bars with the same number of stars above them are statistically equivalent ($\alpha = 0.05$; based on pairwise t -tests performed on regression coefficients; see Table S1, Supplementary material). Labels below each bars indicates the strain name for natural isolates 4A = *SSY1/SSY1*, 4D = *ssy1^{T9}/ssy1^{T9}*, M28 = *SSY1/ssy1^{T9}* and the *SSY1* genotypes for S288c diploid laboratory strains (wt = *SSY1*, KO = *ssy1^{KO}*, T9 = *ssy1^{T9}*). As this figure represents ratios, no error bars are included. See Materials and Methods for a discussion of variance in these growth-rate estimates and their effect on the pairwise t -tests.

the *ssy1^{T9}* in those otherwise isogenic genetic backgrounds. We grew strains in either YPD (rich in amino acids) or in minimal media (MM, lacking amino acids) to test if the presence of a functional *SSY1* has any detectable influence on growth rate in these different environments.

Fig. 4A and 4B show the relative growth rates of M28 and its segregants 4A (*SSY1/SSY1*) and 4D (*ssy1^{T9}/ssy1^{T9}*) in the two media. In Fig. 4(C, F), we report the relative growth rates of S288c-derived diploid strains in these two environments. Given that the gene-expression patterns of S288c *ssy1^{T9}* and S288c *ssy1^{KO}* are essentially the same, we used a diploid knockout to approximate the S288c *ssy1^{T9}/ssy1^{T9}* growth rate. *SSY1* homozygotes and heterozygotes grew faster than *ssy1^{T9}* homozygotes in amino-acid-containing media, regardless of genetic background (Fig. 4A, C). Pairwise t -tests reveal that while strains containing at least one wild-type allele have statistically indistinguishable growth rates, both are significantly greater than *ssy1^{T9}* homozygotes (Table S1, Supplementary material).

While most of the expression differences were masked by growing *ssy1^{T9}/ssy1^{T9}* and *SSY1/SSY1* strains in media without amino acids (see Fig. 1A), our assays revealed that the strains still grow at different rates. Figure 4(B, D) demonstrate that wild-type homozygotes and heterozygotes

grew significantly faster than their *ssy1^{T9}/ssy1^{T9}* homozygote counterparts. The TFL-sensitive 4A segregant from M28 and M28 itself, which are, respectively, *SSY1* homozygous wild type and heterozygous for the frameshift mutation, grew at statistically equivalent rates, while the 4D homozygous *ssy1* mutant segregant grew significantly slower (Fig. 4; Table S1, Supplementary material). Likewise, the growth rates of S288C *SSY1/SSY1* and S288C *SSY1/ssy1^{KO}* were statistically indistinguishable, and both grew significantly faster than S288C *ssy1^{KO}/ssy1^{KO}*. However, S288C *SSY1/ssy1^{T9}* could not be distinguished statistically from any of the other three S288C strains in MM.

Our results imply a higher cost in growth rate due to the *ssy1^{T9}* mutation in minimal media than in YPD. Figure 4 shows that *SSY1* strains grow about 16% faster than their mutant isogenic counterparts compared with only a 5% difference in YPD. This discrepancy was even greater when considering the difference in the progeny of M28, where the wild-type 4A segregant grew 62% faster than the homozygous *ssy1^{T9}* segregant 4D in MM, but only 4% faster in YPD. Put another way, based on growth rate in batch cultures, while the homozygous frameshift resulted in a selective disadvantage of 4–5% in YPD, in MM, it ranged from 13.5% for isogenic derivatives of S288C to

38.2% for the homozygous progeny of M28. This result is somewhat surprising given the similarity of the expression profiles of *SSY1* and *ssy1^{T9}* in MM (Fig. 1), and suggests epistatic effects with one or more alleles in the genetic background.

To assess the statistical power of the growth-rate experiments, we carried out simulations to estimate the probability of rejecting the null hypothesis of equal mean growth rates when there is a stipulated real difference in the means. Since replicate samples in YPD exhibited much smaller standard deviations than those in MM (average standard deviations: YPD = 5.23×10^{-5} ; MM = 2.39×10^{-4} ; Bartlett's K -squared = 92.0377, $P < 2.2 \times 10^{-16}$), we conducted a different set of simulations for each type of medium (see Materials and Methods). In YPD, the simulations indicate that the analyses would reject the null hypothesis of equal mean growth rates 50% of the time if the true difference in means were 2%, and will reject the null hypothesis 80% of the time if the true difference were 3%. In MM, the increased variance among replicates reduces the power of these methods. The analyses would reject the null hypothesis 50% of the time when the true difference in mean growth rate is 10%, and would reject 85% of the time with a true difference in means of 15% (see Fig. S2, Supplementary material).

Discussion

We have isolated a frameshift mutation that segregates in natural populations of wine yeast and causes gene-expression differences at hundreds of loci. This allele explains roughly 50% of the expression differences previously reported among segregants in tetrads from a natural isolate of *Saccharomyces cerevisiae* (M28 in Cavalieri *et al.* 2000). The mutation is an insertion into a mononucleotide repeat in the coding sequence of the gene *SSY1*, which encodes an amino acid sensor. This frameshift mutation disrupts the reading frame of *SSY1*, creating a truncated protein (or a degraded mRNA) that has no detectable residual function. As a major component of the SPS plasma-membrane system for detecting extracellular amino acids, *SSY1* detects and senses amino acids in the environment and triggers the expression of amino acid permeases and transporters, which also feeds into the expression of amino acid biosynthetic enzymes. *SSY1* therefore plays a key role at the interface between the cell's metabolism and its environment and thus in the ecology of *S. cerevisiae*. Our results support previous observations showing that *trans*-acting genetic factors impacting gene expression do not necessarily map to transcription factors or transcriptional regulators (Yvert *et al.* 2003). Interestingly, very few gene deletions in the yeast genome that affect gene expression in *trans* lead to the differential expression of more than 50 genes (Featherstone & Broadie 2002), so sensors of the environment may play an important role in cascading transcriptional effects.

Analysis of gene-expression variation in natural populations has mostly been performed under the framework of quantitative genetics. These analyses traditionally assume that the expression level of each gene is independent and that a large number of genetic factors may contribute to the variation in expression level of each locus. The isolation of a single frameshift mutation that explains genome-wide differences in transcript abundance among natural isolates shows that in some instances very few mutations are necessary to create large and coordinated expression differences. Our results also confirm earlier studies with lower resolution (Brem *et al.* 2002).

Other authors have often noted a positive relation between time of divergence and gene-expression differences (Whitehead & Crawford 2006). Some authors have also suggested that these expression differences accumulate linearly with divergence time (Khaitovich *et al.* 2004). In contrast, our results demonstrate that large differences in gene expression at many loci can be caused by a single mutation. Along with the recent experimental results showing that gene expression levels of different loci are not equally sensitive to spontaneous mutations (Rifkin *et al.* 2003; Denver *et al.* 2005), and that specific properties of gene expression are associated with this differential sensitivity (Landry *et al.* 2007b), our results show that a better understanding of the integration of regulatory networks with other cellular processes is necessary to understand and model the evolution of genome-wide transcriptional profiles.

Several studies have analyzed genome-wide expression profiling with the goal of disentangling the role of various evolutionary forces on the transcriptome and their functional mechanisms. Comparisons of gene-expression variation across functional categories and network properties indirectly suggested that gene-expression polymorphism is constrained by the gene's essentiality (Landry *et al.* 2006a) as well as its by position in the protein interaction network (Lemos *et al.* 2005). Here we used a different, more direct approach. We first isolated a mutation that contributes to gene-expression differences and studied the effects of this mutation on growth rate. Using the selection coefficients estimated from our growth-rate analyses along with reasonable approximations of mononucleotide repeat mutation rates, we can estimate the frequency of the *ssy1^{T9}* allele in natural *S. cerevisiae* populations. Homopolymeric repeats, such as the eight thymidylates in *SSY1*, mutate at high rates. Previous work has estimated mutation rates equal to if not exceeding rates for microsatellite loci. Lynch *et al.* (in press) used complete genome sequencing and estimated mutation rates at homopolymeric runs to be between 10^{-5} and 10^{-6} per cell division (but see Gragg *et al.* 2002 for homopolymeric runs in plasmids). As the frameshift allele presumably exists at low frequencies, most copies will occur in heterozygotes. Given the limitations of our experiments and statistical power in detecting growth rate

differences, *ssy1^{T9}* may have a small selective effect in heterozygotes (i.e. potentially as large as 2% in YPD and 10% in MM, see Results). Additionally, natural populations of yeast rarely undergo sexual reproduction, indicating that most selection on this low frequency allele would occur in heterozygotes (Ruderfer *et al.* 2006). Under these conditions, population-genetics theory (Hartl & Clark 2007) predicts an equilibrium frequency of the mutant allele as $q = u/hs$, where u is the forward mutation rate, s is the reduction in fitness in homozygous genotypes (here estimated as 0.047 in YPD), and h is the degree of dominance. Assuming a typical value $h = 0.02$ (Simmons & Crow 1977), then q would range from 0.01 to 0.001, or the frequency of heterozygous genotypes from 1/50 to 1/500. A smaller degree of dominance or smaller selection coefficient when homozygous would predict higher natural frequencies for this allele. In any case, *ssy1^{T9}* is predicted to be about as frequent as many human genetic disease alleles with between one in 50–500 individuals being heterozygous (Reich & Lander 2001).

By identifying the simple genetic basis for half of the gene-expression differences among M28 progeny, we have demonstrated that single-nucleotide mutations can have highly pleiotropic effects on gene expression. While the *ssy1^{T9}* mutation acts in *trans* to affect the expression level of a hundred or more genes, it is not immediately associated with transcription factors. If *ssy1^{T9}* is any indication, recessive mutations, especially those associated with unstable repeats, may persist at low but significant frequencies. While rarely unmasked, alleles such as *ssy1^{T9}* can generate a relatively large amount of phenotypic diversity from small amounts of genotypic diversity.

Acknowledgements

We thank Bernardo Lemos and Pierre Fontanillas for helpful discussions regarding microarray analysis and for their critical review of this manuscript. We also thank Kevin Verstrepen and the FAS Center for Systems Biology for the use of their facilities and Fred Winston and Jeffrey Townsend for providing yeast strains. We are also grateful to two anonymous reviewers for critical review of the manuscript. KMB is supported by a National Science Foundation Graduate Research Fellowship. CRL was supported for this research by the Natural Sciences and Engineering Research Council of Canada (NSERC), Fonds québécois de la recherche sur la nature et les technologies (FQRNT) and the Frank Knox Memorial Foundation.

References

- Aa E, Townsend JP, Adams RI, Nielsen KM, Taylor JW (2006) Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, **6**, 702–716.
- Adams A, Gottschling DE, Kaiser CA, Stearns T (1997) *Methods in Yeast Genetics: a Cold Spring Harbor Laboratory Course Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Aubin-Horth N, Landry CR, Letcher BH, Hofmann HA (2005) Alternative life histories shape brain gene expression profiles in males of the same population. *Proceedings of the Royal Society, Series B: Biological Sciences*, **272**, 1655–1662.
- Ben-Ari G, Zenvirth D, Sherman A *et al.* (2006) Four linked genes participate in controlling sporulation efficiency in budding yeast. *Plos Genetics*, **2**, 1815–1823.
- Bernard F, Andre B (2001) Genetic analysis of the signalling pathway activated by external amino acids in *Saccharomyces cerevisiae*. *Molecular Microbiology*, **41**, 489–502.
- Brachmann CB, Davies A, Cost GJ *et al.* (1998) Designer Deletion Strains derived from *Saccharomyces cerevisiae* S288C: a Useful set of Strains and Plasmids for PCR-mediated Gene Disruption and Other applications. *Yeast*, **14**, 115–132.
- Bradshaw Jr HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176–178.
- Bradshaw Jr HD, Wilbert SM, Otto KG, Schemske DW (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers. *Nature*, **376**, 762–765.
- Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences, USA*, **102**, 1572–1577.
- Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.
- Cavaliere D, Barberio C, Casalone E *et al.* (1998) Genetic and molecular diversity in *Saccharomyces cerevisiae* natural populations. *Food Technology and Biotechnology*, **36**, 45–50.
- Cavaliere D, Townsend JP, Hartl DL (2000) Manifold anomalies in gene expression in vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proceedings of the National Academy of Sciences, USA*, **97**, 12369–12374.
- Clark TA, Townsend JP (2007) Quantifying variation in gene expression. *Molecular Ecology*, **16**, 2613–2616.
- Denver DR, Morris K, Streelman JT *et al.* (2005) The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nature Genetics*, **37**, 544–548.
- Eckert-Boulet N, Nielsen PS, Friis C *et al.* (2004) Transcriptional profiling of extracellular amino acid sensing in *Saccharomyces cerevisiae* and the role of Stp1p and Stp2p. *Yeast*, **21**, 635–648.
- Erickson DL, Fenster CB, Stenoien HK, Price D (2004) Quantitative trait locus analyses and the study of evolutionary process. *Molecular Ecology*, **13**, 2505–2522.
- Fay JC, McCullough HL, Sniegowski PD, Eisen MB (2004) Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biology*, **5**, R26.
- Featherstone DE, Broadie K (2002) Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *Bioessays*, **24**, 267–274.
- Forsberg H, Ljungdahl PO (2001) Genetic and biochemical analysis of the yeast plasma membrane Ssy1p-Ptr3p-Ssy5p sensor of extracellular amino acids. *Molecular and Cellular Biology*, **21**, 814–826.
- Forsberg H, Gilstring CF, Zargari A, Martinez P, Ljungdahl PO (2001) The role of the yeast plasma membrane SPS nutrient sensor in the metabolic response to extracellular amino acids. *Molecular Microbiology*, **42**, 215–228.
- Gibson G, Riley-Berger R, Harshman L *et al.* (2004) Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics*, **167**, 1791–1799.

- Gragg H, Harfe BD, Jinks-Robertson S (2002) Base composition of mononucleotide runs affect DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, **22**, 8756–8762.
- Hartl DL, Clark AG (2007) *Principles of Population Genetics*. Sinauer Associates Inc. Publishers, Sunderland, MA.
- Hsieh WP, Chu TM, Wolfinger RD, Gibson G (2003) Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics*, **165**, 747–757.
- Jin W, Riley RM, Wolfinger RD *et al.* (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics*, **29**, 389–395.
- Jordan IK, Marino-Ramirez L, Koonin EV (2005) Evolutionary significance of gene expression divergence. *Gene*, **345**, 119–126.
- Joseph SB, Hall DW (2004) Spontaneous Mutations in Diploid *Saccharomyces cerevisiae*: More Beneficial Than Expected. *Genetics*, **168**, 1817–1825.
- Khaitovich P, Weiss G, Lachmann M *et al.* (2004) A neutral model of transcriptome evolution. *Plos Biology*, **2**, 682–689.
- Klasson H, Fink GR, Ljungdahl PO (1999) Ssy1p and Ptr3p are plasma membrane components of a yeast system that senses extracellular amino acids. *Molecular and Cellular Biology*, **19**, 5405–5416.
- Landry CR, Oh J, Hartl DL, Cavalieri D (2006a) Genome-wide scan reveals that genetic variation for transcriptional plasticity in yeast is biased towards multi-copy and dispensable genes. *Gene*, **366**, 343–351.
- Landry CR, Townsend JP, Hartl DL, Cavalieri D (2006b) Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Molecular Ecology*, **15**, 575–591.
- Landry CR, Castillo-Davis CI, Ogura A, Liu JS, Hartl DL (2007a) Systems-level analysis and evolution of the phototransduction network in *Drosophila*. *Proceedings of the National Academy of Sciences, USA*, **104**, 3283–3288.
- Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL (2007b) Genetic properties influencing the evolvability of gene expression. *Science*, **317**, 118–121.
- Lemos B, Meiklejohn CD, Caceres M, Hartl DL (2005) Rates of divergence in gene expression profiles of primates, mice, and flies stabilizing selection and variability among functional categories. *Evolution*, **59**, 126–137.
- Lemos B, Landry CR, Fontanillas P *et al.* (2007) Evolution of genomic expression. In: *Evolutionary Genomics and Proteomics* (eds Pagel M, Pomiankowski A). Sinauer Associates. Inc. Publishers, Sunderland, MA.
- Longtine MS, McKenzie A, Demarini DJ *et al.* (1998) Additional Modules for Versatile and Economical PCR-based Gene Deletion and modification in *Saccharomyces cerevisiae*. *Yeast*, **14**, 953–961.
- Lynch M, Sung W, Morris K *et al.* (in press) A genome-wide view of the spectrum of spontaneous mutations in Yeast. *Proceedings of the National Academy of Sciences, USA*.
- Mortimer RK (2000) Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Research*, **10**, 403–409.
- Mortimer R, Polsinelli M (1999) On the origins of wine yeast. *Research in Microbiology*, **150**, 199–204.
- Poulsen P, Wu B, Gaber RF, Kielland-Brandt MC (2005) Constitutive signal transduction by mutant Ssy5p and Ptr3p components of the SPS amino acid sensor system in *Saccharomyces cerevisiae*. *Eukaryotic Cell*, **4**, 1116–1124.
- Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL (2003) Sex-Dependent Gene Expression and the evolution of the *Drosophila* transcriptome. *Science*, **300**, 1742–1745.
- Reich DE, Lander ES (2001) On the allelic spectrum of human disease. *Trends in Genetics*, **17**, 502–510.
- Rifkin SA, Kim J, White KP (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genetics*, **33**, 138–144.
- Ron M, Weller JI (2007) From QTL to QTN identification in livestock – winning by points rather than knock-out: a review. *Animal Genetics*, **38**, 429–439.
- Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L (2006) Population genomic analysis of outcrossing and recombination in yeast. *Nature Genetics*, **38**, 1077–1081.
- Sikorski RS, Hieter P (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*, **122**, 19–27.
- Simmons MJ, Crow JF (1977) Mutations affecting fitness in *Drosophila* populations. *Annual Review of Genetics*, **11**, 49–78.
- Smyth GK, Speed TP (2003) Normalization of cDNA microarray data. *Methods*, **31**, 265–273.
- Steinmetz LM, Sinha H, Richards DR *et al.* (2002) Dissecting the architecture of a quantitative trait locus in yeast. *Nature*, **416**, 326–330.
- Townsend JP (2004) Reduction of large and small differences in gene expression using models for the Bayesian analysis of gene expression levels and spotted DNA microarrays. *BMC Bioinformatics*, **5**, 54–66.
- Townsend JP, Hartl DL (2002) Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple treatments or samples. *Genome Biology*, **3**, research0071.1-0071.16.
- Townsend JP, Cavalieri D, Hartl DL (2003) Population genetic variation in genome-wide gene expression. *Molecular Biology and Evolution*, **20**, 955–963.
- Whitehead A, Crawford DL (2006) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, **15**, 1197–1211.
- Winston F, Dollard C, Ricupero-Hovasse SL (1995) Construction of a set of congruent *Saccharomyces cerevisiae* that are isogenic to S288C. *Yeast*, **11**, 53–55.
- Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart BJ (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnology*, **15**, 1359–1367.
- Yvert G, Brem RB, Whittle J *et al.* (2003) Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genetics*, **35**, 57–64.

Kyle M. Brown is a PhD candidate with Daniel L. Hartl. He is interested in evolutionary genetics and his research focuses on the evolution of gene expression and the role of epistasis in the evolution of antibiotic resistance. Christian Landry is a NSERC and CIHR postdoctoral fellow at Université de Montreal where he studies the functional evolution of gene and protein networks. Daniel L. Hartl's research interests include evolutionary genomics, molecular evolution and population genetics. Duccio Cavalieri's research focuses on investigating how genetic diversity is generated and maintained and how it affects the behaviour of yeast populations in laboratory and natural settings.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Microarray experimental design. We compared the expression profile of the wild-type S228c strain with that of isogenic strains in which we knocked-out the *SSY1* (*ssy1^{ko}*) or introduced the T9 frameshift mutation (*ssy1^{T9}*). The three strains were compared to each other directly on the microarrays and in the two possible dyes orientation (dye-swap) for a total of 6 hybridizations.

Fig. S2 Simulation results examining *t*-test statistical power. (A) Frequency of null hypotheses rejected using a given level of the absolute difference in mean growth rate. (B) The same simulation data plotted as a function of the percent difference between the

two mean growth rates. See Materials and Methods for complete description of simulations.

Table S1 Comparisons of growth rates among isogenic or nearly isogenic strains in different media. *P*-values from pairwise Student's *t*-tests are reported. (wt = *SSY1*, KO = *ssy1^{KO}*, T9 = *ssy1^{T9}*). *P*-values less than 0.05 are starred and in bold

This material is available as part of the online article from:
<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2008.03765.x>
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.