

The Genomic Architecture of Interactions Between Natural Genetic Polymorphisms and Environments in Yeast Growth

Xinzhu Wei and Jianzhi Zhang¹

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109

ABSTRACT Gene-environment interaction (G×E) refers to the phenomenon that the same mutation has different phenotypic effects in different environments. Although quantitative trait loci (QTLs) exhibiting G×E have been reported, little is known about the general properties of G×E, and those of its underlying QTLs. Here, we use the genotypes of 1005 segregants from a cross between two *Saccharomyces cerevisiae* strains, and the growth rates of these segregants in 47 environments, to identify growth rate QTLs (gQTLs) in each environment, and QTLs that have different growth effects in each pair of environments (g×eQTLs). The average number of g×eQTLs identified between two environments is 0.58 times the number of unique gQTLs identified in these environments, revealing a high abundance of G×E. Eighty-seven percent of g×eQTLs belong to gQTLs, supporting the practice of identifying g×eQTLs from gQTLs. Most g×eQTLs identified from gQTLs have concordant effects between environments, but, as the effect size of a mutation in one environment enlarges, the probability of antagonism in the other environment increases. Antagonistic g×eQTLs are enriched in dissimilar environments. Relative to gQTLs, g×eQTLs tend to occur at intronic and synonymous sites. The gene ontology (GO) distributions of gQTLs and g×eQTLs are significantly different, as are those of antagonistic and concordant g×eQTLs. Simulations based on the yeast data showed that ignoring G×E causes substantial missing heritability. Together, our findings reveal the genomic architecture of G×E in yeast growth, and demonstrate the importance of G×E in explaining phenotypic variation and missing heritability.

KEYWORDS antagonism; missing heritability; pleiotropy; QTL mapping; *Saccharomyces cerevisiae*

GENE-ENVIRONMENT interaction (G×E) refers to the observation that the same mutation has different phenotypic effects on a trait in different environments (Ottman 1996). G×E is believed to be ubiquitous among all organisms, and has long been studied in domestic animals and plants, genetic model organisms, and humans. In humans, G×E has been implicated in cancer (Thorgeirsson *et al.* 2008), inflammatory disorder (Chamaillard *et al.* 2003), immune system diseases (Padyukov *et al.* 2004), and mental disorders (Risch *et al.* 2009; Byrd and Manuck 2014; Luck *et al.* 2014). Investigating G×E can help identify the causal pathways of a trait (Gagneur *et al.* 2013), dissect genetic trade-offs (Qian *et al.* 2012), understand environmental adaptations

(Ostrowski *et al.* 2005), and reveal a potential cause of “missing heritability” (Manolio *et al.* 2009; Eichler *et al.* 2010).

G×E studies can be generally divided into two types on the basis of the approach used: forward genetics and reverse genetics. In forward genetics, genes or quantitative trait loci (QTLs) that show significantly different phenotypic effects in different environments are identified via linkage or association mapping. In reverse genetics, a mutant carrying a known mutation such as a gene deletion or a point mutation is compared with the wild-type for the trait of interest under two environments, and G×E is detected when the mutational effect on the trait differs significantly in the two environments. For example, Qian and colleagues measured the fitness effects of single gene deletions in yeast for nearly 5000 nonessential genes in six different environments, and identified many antagonistic G×E cases where deleting a gene is deleterious in one environment but beneficial in another (Qian *et al.* 2012). Although such systematic reverse genetic studies can provide a broad picture of G×E, to date they are limited to gene deletions (Dudley *et al.* 2005; Brown *et al.* 2006; Hillenmeyer *et al.* 2008; Qian *et al.* 2012), which

Copyright © 2017 by the Genetics Society of America
doi: 10.1534/genetics.116.195487

Manuscript received August 31, 2016; accepted for publication November 22, 2016; published Early Online November 28, 2016.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.195487/-/DC1.

¹Corresponding author: Department of Ecology and Evolutionary Biology, University of Michigan, 1075 Natural Science Bldg., 830 North University Ave., Ann Arbor, MI 48109. E-mail: jjanzhi@umich.edu

constitute a special group of mutations. In theory, the reverse genetic approach can also be applied to all natural genetic polymorphisms, but studies of this sort are universally small in scale (Ostrowski *et al.* 2005; Gerke *et al.* 2010; Dillon *et al.* 2016), and thus do not offer an overview of G×E for natural genetic polymorphisms. By contrast, large forward genetic analysis in principle allows deciphering general properties of G×E for natural genetic variants.

Many recent forward genetic studies of G×E in humans are driven by the idea of personalized medicine, and focus on finding candidate genes and environmental factors that interact in influencing disease, drug response, or behavior (Caspi *et al.* 2002, 2005; Hood *et al.* 2004; Kendler *et al.* 2012; Byrd and Manuck 2014; Luck *et al.* 2014). Although a number of genes have been reported to interact with environmental factors, the reproducibility of these genome-wide association study (GWAS) results tends to be low (Hunter 2005; Duncan and Keller 2011), and one likely reason is that environmental factors are hard to control in human studies. The power to detect genetic variants that interact with environments is generally lower than the power to detect genetic variants that have effects in one environment. Furthermore, the detection of interaction is affected by how interaction is measured (Duncan and Keller 2011), because the null hypothesis of no interaction may be based on an additivity or multiplicity assumption. That is, if the phenotypes of two genotypes are A1 and B1 in environment 1 and A2 and B2 in environment 2, respectively, the null hypothesis of no G×E under additivity is $A1 - B1 = A2 - B2$, whereas that under multiplicity is $A1/B1 = A2/B2$. In model organisms such as the mouse *Mus musculus* and fly *Drosophila melanogaster*, recombinant inbred lines established from a cross between two parental lines are typically used to identify G×E QTLs via linkage mapping (Fry *et al.* 1998; Ungerer *et al.* 2003; Li *et al.* 2006; Flint and Mackay 2009; Gerke *et al.* 2010; El-Soda *et al.* 2014; Matsui and Ehrenreich 2016). Generally speaking, environments are better controlled, detection power is higher, and the detected interactions are more readily verifiable in model organism studies, compared with human studies.

Although the abundance of G×E has been demonstrated in various model organisms, there has been no systematic study of the genomic and functional distributions of G×E QTLs. Furthermore, it is unknown whether G×E is mostly antagonistic (*i.e.*, the same allele has opposite phenotypic effects in two environments) or concordant among natural genetic polymorphisms. It is also unclear how much ignoring G×E impacts the identification of QTLs underlying natural phenotypic variations among individuals that cannot possibly have identical environments. Methodologically, some human studies identify G×E by directly testing if genes with known effects in one environment have different effects in another environment (Caspi *et al.* 2003), instead of testing all genetic variants by GWAS. Although the former approach has been criticized to have publication bias, low statistical power, and high false discovery rates when compared with GWAS (Duncan and Keller 2011), some authors consider it to be

more replicable and superior for finding causal genes (Moffitt *et al.* 2005; Uher 2014). Which of the two methods performs better depends on the probability that an influential mutation in one environment has a different effect in another environment. It also depends on the probability that a G×E QTL between two environments has detectable effects in at least one of the environments. But neither of these probabilities is currently known. Here, we address all these questions using a recently published dataset of the budding yeast *Saccharomyces cerevisiae*, which includes the genome sequences, and the growth rates in 47 environments, of 1005 haploid segregants produced by the F1 resulting from a cross between strains BY and RM (Bloom *et al.* 2013). BY is derived from the commonly used laboratory strain S288c, whereas RM is derived from the vineyard strain RM11-1a. The 47 growth environments varied in temperature, pH, carbon source, metal ions, and small molecules (Bloom *et al.* 2013). The growth rate of each segregant was measured by the mean end-point colony radius on agar plates. Although a more recently published dataset (Bloom *et al.* 2015) contained 4390 segregants from the same F1, only 21 environments were examined. We thus focused on the earlier data, which include more environments, and hence better suit the study of G×E. We analyzed the later data (Bloom *et al.* 2015), only to verify the key findings from the earlier data. Note that several yeast studies mapped growth rate QTLs in each of an array of environments (Cubillos *et al.* 2011; Ehrenreich *et al.* 2012; Bloom *et al.* 2013; Wilkening *et al.* 2014), or mapped plasticity QTLs across environments (Yadav *et al.* 2016), but these studies either treated growth rates in different environments as different traits, or treated growth rate variance among environments as a phenotypic trait. Hence, yeast G×E in growth rate has not been studied.

Materials and Methods

Genotype and phenotype data

We acquired from the Kruglyak laboratory the genotype data of 1040 segregants from a cross between the BY and RM strains of *S. cerevisiae*, including a total of 28,220 single nucleotide polymorphisms (SNPs) mapped to the reference genome sequence R64-1-1 (Bloom *et al.* 2013). We similarly obtained the average end-point colony radius of each segregant in each of the 47 environments (Bloom *et al.* 2013). After requiring each segregant to have both genotype data and phenotype data in at least one environment, we retained 1005 qualified segregants for subsequent analysis. Narrow-sense heritability data were from the Supplemental Materials of the original publication (Bloom *et al.* 2013). We also acquired the genotype and phenotype data from a follow-up study (Bloom *et al.* 2015), where the growth rates of 4390 segregants from the same cross were similarly measured in 21 of the original 47 environments. We downloaded the cDNA sequences, genome annotations, GO terms, and GO domains from Ensembl biomart for reference R64-1-1, and used Matlab scripts for all enrichment tests.

Mapping growth rate QTLs (gQTLs) in an environment

We started the first round of gQTL mapping using the filtered growth rates as the phenotype. The filtered growth rate of a segregant is its colony radius after 48 hr growth on agar plates averaged between two replicates, followed by a series of data filtering and correction by the original authors (Bloom *et al.* 2013). Given an environment, for each SNP, we compared the growth rates between the two groups of segregants that carry the alternative alleles, using a *t*-test. We converted *P*-values to *Q*-values (Storey and Tibshirani 2003). A stringent *Q*-value of 0.005 was used as the cutoff for statistical significance, on the basis of the simulation described below. On each chromosome, we chose the SNP with the lowest *Q*-value. Sometimes, a chromosome carried multiple SNPs with exactly the same minimal *Q*-values; these were always adjacent SNPs (*i.e.*, with no intervening SNP), and the middle SNP was chosen. We combined all chosen SNPs from all chromosomes to fit the linear model $Y = \beta_0 + \beta X + \varepsilon$, where *Y* is a vector of the growth rates of all segregants, β_0 is the fitted population mean growth rate, β is a vector of gQTL effect sizes, ε is an error vector, and *X* is a matrix of genotypes (number of segregants \times number of gQTLs). If the allele at a SNP is from BY, the corresponding element in *X* is -1 ; otherwise, it is 1. We estimated β , growth rate residuals, and *t*-statistics from regression using the embedded Matlab function LinearModel. A SNP is removed if its contribution in the linear model is not significant at $P = 0.05$ by a *t*-test. We then used all remaining SNPs to fit a linear model and calculated the growth rate residuals.

We started the second round of gQTL mapping using the growth rate residuals as phenotypes, following the procedure described above. We then combined the SNPs identified from the first two cycles to fit a linear model, removed SNPs with insignificant contribution to the linear model, and calculated growth rate residuals using the remaining SNPs. This process was repeated until no more SNPs were added in a cycle of gQTL mapping. In all environments, four or fewer cycles were needed. That is, each chromosome has at most three gQTLs identified in an environment.

Mapping growth rate by environment interaction QTLs (g \times eQTLs) in each pair of environments

The 47 environments form 1081 pairs. We first used the identified gQTLs to test G \times E (class I g \times eQTLs). That is, for a given environment pair, and a gQTL identified from one or both of these two environments, we used a genotype's growth rate difference between the two environments as its phenotype, and then used a *t*-test to compare the phenotypes of the groups of genotypes with alternative alleles at the gQTL. $P = 0.05$ from a *t*-test was used to determine whether significant G \times E is present for the gQTL; simulation results suggested no need for multiple-testing correction here. Given that on average only 10.3 gQTLs were mapped per environment, we assumed that any two gQTLs that are identified from different environments and lie within 7500 nucleotides of each other (corresponding to the average distance

spanned by ~ 4 genes) have the same underlying causal genetic variant. In such cases, we tested the middle SNP between the two gQTLs for G \times E. The justification of the above assumption is as follows: if the gQTLs from two environments are independent from each other, and are randomly distributed across the genome, the probability that a gQTL identified in one environment is within 7500 nucleotides of a gQTL identified in the other environment is 1.3%. In fact, an average of 11.0% of gQTLs identified in one environment are within 7500 nucleotides from a gQTL identified in the other environment, suggesting that the vast majority of gQTLs within 7500 nucleotides from each other are not independent, but share the same causal mutation.

For each environment pair, we also mapped class II g \times eQTLs by considering all SNPs. The method used was the same as mapping gQTLs in an environment, except that growth rate differences between two environments instead of growth rates in one environment were used as phenotypes. We first calculated the difference in end-point colony radius between the two environments for each segregant that has the colony radius measures in both environments, and then followed the same procedure as gQTL mapping to identify class II g \times eQTLs. We similarly terminated the search when no more SNPs were added to the model. A *Q*-value of 0.005 was used as the cutoff for statistical significance, on the basis of the simulation described below. We counted class II g \times eQTLs mapped on chromosomes with no gQTL from either environment. We focused on these chromosomes, because it would otherwise be unclear if class I and class II g \times eQTLs reflect the same causal SNPs, owing to strong linkage of SNPs within a chromosome. From the number of class II g \times eQTLs on these chromosomes, we extrapolated the number of class II g \times eQTLs in the entire genome on the basis of the relative sizes of the chromosomes, under the assumption that class II g \times eQTLs are evenly distributed across the genome. Extrapolated class II g \times eQTLs were used only to estimate the g \times eQTLs missed by class I g \times eQTL mapping.

Computer simulation for determining the *Q*-value cutoff

We converted *P*-values to *Q*-values according to the method of Storey and Tibshirani (2003), because it is in theory ~ 1000 times faster than obtaining *Q*-values from the permutation test used in the original analysis of this dataset (Bloom *et al.* 2013). We used computer simulation to compare the performance of our method with the one previously used (Bloom *et al.* 2013) in order to choose a proper *Q*-value cutoff. To save computational time, we simulated three chromosomes instead of all 16 chromosomes in the yeast genome, using parameters appropriate for average-size yeast chromosomes. Each simulated chromosome carried 1500 SNPs, and two recombination events were randomly allocated per chromosome in each segregant on the basis of 90.5 crossovers per yeast meiosis (Mancera *et al.* 2008). We randomly assigned three SNPs that are >30 SNPs away from one another to be gQTLs. Phenotypic noise is simulated using the standard normal

distribution. In the first simulation, each of the three gQTLs has an effect size = 1, and one of the two alleles at a gQTL is randomly picked to be the fitter allele. The narrow-sense heritability $h^2 = 3 \times 1^2 / (3 \times 1^2 + 1) = 0.75$. In the second and third simulations, we used the effect size of 0.75 and 0.5, respectively, corresponding to $h^2 = 0.63$ and 0.43, respectively. These h^2 values match approximately the observed h^2 values in our data. Each simulation generated 1000 segregants. We then mapped gQTLs using different Storey and Tibshirani Q -value cutoffs (0.05, 0.02, 0.01, 0.005, 0.002, and 0.001) in our method, and compared our results with those of Bloom *et al.* (2013) that were based on the permutation Q -value of 0.05. The false discovery and false negative rates were estimated for both methods. We found that, for both methods, the false discovery rates were greater than the rate the Q -values suggested, but false negative rates were negligibly small. The false discovery rates of our method under Q -values of 0.01 and 0.005 were comparable to those derived from the method of Bloom *et al.* (2003). We thus chose the more stringent Q -value cutoff of 0.005 in our mapping.

We also simulated an environment pair with the parameters used above. That is, three gQTLs existed in each environment, but they had no effect in the other environment. We then mapped gQTLs with a Q -value cutoff of 0.005, followed by class I g×eQTLs mapping with a P -value cutoff of 0.05. The results obtained are presented in Supplemental Material, Table S1. Because our detection of gQTLs had very low false negative rates (Table S1), we were not able to study the performance of identifying class II g×eQTLs by our simulation. One type of gQTLs not considered in the above simulation is those that have the same effects in two environments. Such gQTLs could be erroneously identified as g×eQTLs. To examine the probability of this error, we simulated three gQTLs with the same effects in two environments. We found that this type of false positive error hardly increases the overall false discovery rate of g×eQTLs, and therefore we did not include it in Table S1.

Data availability

All yeast genotype and phenotype data were previously published (Bloom *et al.* 2013, 2015). The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

Results

Identification of QTLs that interact with environments

Because we aimed to identify G×E in all $47 \times 46/2 = 1081$ environment pairs, a computationally efficient mapping method was needed. To this end, we developed a customized rapid QTL mapping method with a false discovery rate comparable to that of a previous method (Bloom *et al.* 2013), and validated its performance by computer simulation (Table S1; see *Materials and Methods*). With the new method, we first identified QTLs underlying the among-segregant growth rate

variation in each environment using the genotype and phenotype data of the 1005 segregants. The identified QTLs are denoted as gQTLs, where “g” stands for growth rate. We were able to identify gQTLs in 45 of the 47 environments (File S1). The number of gQTLs ranges from 0 to 22 across the 47 environments, with the mean = 10.3. We calculated the similarity between two environments by the across-segregant rank correlation between growth rates in the two environments. The higher the similarity between two environments, the smaller the difference in the number of gQTLs mapped in these environments (Spearman’s $\rho = -0.26$, $P < 10^{-17}$).

We then attempted to identify loci exhibiting G×E (g×eQTLs) for each of the 1081 environment pairs. We used the gQTLs identified from each of the two environments under consideration, and tested if a gQTL has significantly different effects in the two environments. This approach is based on the premise that a g×eQTL should have a phenotypic effect (though not necessarily significant) in at least one of the two environments compared. We used this approach rather than directly testing each SNP for G×E, because the former is expected to have a higher signal to noise ratio such that the identified g×eQTLs are more likely to be genuine. This expectation was confirmed by computer simulation. Specifically, the false discovery rate was lower and the identified g×eQTLs were closer to the causal SNPs when comparing our approach with directly testing all SNPs for G×E (Table S1 and Table S2; see *Materials and Methods*). Nevertheless, if the phenotypic effects of a locus in two environments are both small, the locus may be detected as a gQTL in neither environment. Thus, even if the locus has a significant G×E effect, it may be missed by our approach. To rectify this problem, we also directly mapped G×E for all SNPs, but considered only those that are on chromosomes where no gQTL in the relevant environments was found by the first approach (see *Materials and Methods*). We focused on these chromosomes because it would otherwise be unclear if g×eQTLs identified by the two approaches reflect the same causal SNPs, owing to strong linkage of SNPs within a chromosome, and because the performance in detecting g×eQTLs is better for the first approach than the second approach. The g×eQTLs identified by the two approaches are referred to as class I and class II g×eQTLs, respectively. Considering the total length of chromosomes where class II g×eQTLs are considered and the total length of all yeast chromosomes, we extrapolated the expected number of class II g×eQTLs for the entire genome from that of the considered ones. They are respectively referred to as the extrapolated number and the observed number of class II g×eQTLs.

Class I g×eQTLs outnumber class II g×eQTLs

As an example, let us examine the gQTLs identified under two environments: hydrogen peroxide (HydPer) medium and indoleacetic acid (IndAci) medium, as well as the g×eQTLs identified for this pair of environments (Figure 1A). Nine gQTLs were identified in HydPer and 13 in IndAci. The RM allele is fitter than the BY allele at 13 gQTLs, while the opposite is true at the other nine gQTLs. We identified eight

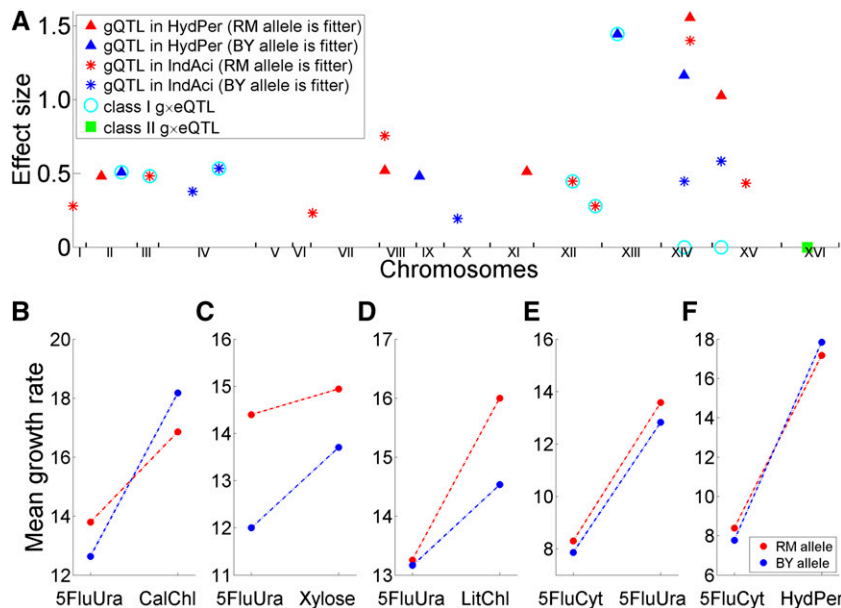


Figure 1 Examples of gQTLs and g×eQTLs. (A) Genomic distributions of detected gQTLs in HydPer and IndAci, and g×eQTLs between the two environments. The effect size of a gQTL under the environment where it is identified is shown on the Y-axis, while its genomic position is shown on the X-axis. A class I g×eQTL is circled at the triangle if it is a gQTL only in HydPer, and circled at the star if it is a gQTL only in IndAci, but is circled on the X-axis if it is a gQTL in both environments. Observed class II g×eQTLs are indicated on the X-axis. (B–F) Mean growth rates of segregants carrying the two alternative alleles at various gQTLs or g×eQTLs. SE are too small to see. (B) A class I antagonistic g×eQTL that is a gQTL (SNP: 24637) in both 5FluUra and CalChl. (C) A class I concordant g×eQTL (SNP: 24651) that is a gQTL in both 5FluUra and Xylose. (D) A class I g×eQTL that is a gQTL (SNP: 4821) in LitChl but not 5FluUra. (E) A gQTL (SNP: 2277) in 5FluUra that does not show significant G×E. (F) A class II antagonistic g×eQTL (SNP: 3512), which is a gQTL in neither 5FluUra nor HydPer.

class I g×eQTLs and observed one class II g×eQTL. Some clear examples of various types of G×E, not necessarily from the above environment pair, are shown in Figure 1, B–F. In these examples, g×eQTLs are found on chromosomes with, at most, one mapped gQTL, so the difference in mean growth rate between genotypes of alternative alleles likely represents primarily the g×eQTL effect without influences from linked gQTLs. For instance, Figure 1B shows a gQTL identified from both 5-fluorouracil (5FluUra) and calcium chloride (CalChl) but with alternative fitter alleles. Not surprisingly, it is a class I antagonistic g×eQTL (*i.e.*, the effects of an allele in the two environments are of opposite direction). Figure 1C shows a gQTL identified from both 5FluUra and Xylose. Although the RM allele is the fitter allele in both environments, the effect size differs; this gQTL is thus a concordant class I g×eQTL (*i.e.*, the effects of an allele in the two environments are of the same direction). Figure 1D shows a gQTL identified in only one of the two environments [lithium chloride (LitChl)], and it is a class I antagonistic g×eQTL. Figure 1E shows a gQTL identified in 5FluUra but not in 5-fluorocytosine (5FluCyt), and it does not have a significant G×E effect between the two environments. Figure 1F shows a locus that is not a gQTL in either 5FluCyt or HydPer, but is a class II g×eQTL.

The numbers of gQTLs, class I g×eQTLs, and observed class II g×eQTLs found in each 3 cM (7500-nucleotide or four-gene) segment along the yeast genome for all environments and environment pairs considered are presented in Figure 2. The total number of gQTLs identified from 47 environments in a 3 cM segment ranges from 0 to 17 (Figure 2A). The number of class I g×eQTLs from all environment pairs in a 3 cM segment ranges from 0 to 374 (Figure 2B), while the corresponding number of observed class II g×eQTLs ranges from 0 to 13 (Figure 2C). The numbers of gQTLs and class I g×eQTLs across 3 cM segments are highly correlated (Pearson's $r = 0.901$, $P < 10^{-250}$), while those of gQTLs

and class II g×eQTLs are distinct ($r = 0.011$, $P = 0.67$) (Figure 2). On average, there are 9.2 class I g×eQTLs, but only 0.37 observed class II g×eQTLs per environment pair, the former being significantly greater than the latter ($P < 10^{-250}$). The same trend was observed when extrapolated instead of observed class II g×eQTLs were considered ($P < 10^{-161}$). We tested three genes (*HAP1*, *MKT1*, and *IRA2*) that accounted for much of the deviation from null in a previous gene expression G×E study of the same strain pair between glucose and ethanol environments (Smith and Kruglyak 2008). Interestingly, these genes locate in 3 cM segments frequently harboring gQTLs and class I g×eQTLs in our study as well. Specifically, *IRA2*, encoding a GTPase-activating protein that modulates the metaphase to anaphase transition during yeast mitosis (Luo *et al.* 2014), overlaps with the segment that has the highest numbers of gQTLs and class I g×eQTLs among all segments (Figure 2). All class I g×eQTLs mapped are listed in File S2.

For each environment pair, we computed the ratio between the number of class I g×eQTLs and the total number of unique gQTLs (*i.e.*, shared gQTLs between the environments are counted only once) identified (Figure 3A). The ratio averages 0.45 across all environment pairs. Many human studies tested G×E by considering candidate genes that are previously known or predicted to have effects in at least one of the environments compared (Duncan and Keller 2011). Across environment pairs in our data, on average 87% of all g×eQTLs (*i.e.*, class I g×eQTLs plus extrapolated class II g×eQTLs) are class I (Figure 3B), supporting the validity of this practice. The number of g×eQTLs for a pair of environments is, on average, 0.58 times the total number of unique gQTLs in these environments (Figure 3C), indicating the high abundance of G×E.

Antagonistic G×E is uncommon

Previous case studies in *Escherichia coli*, *D. melanogaster*, and *Arabidopsis thaliana* suggested the scarcity of antagonistic

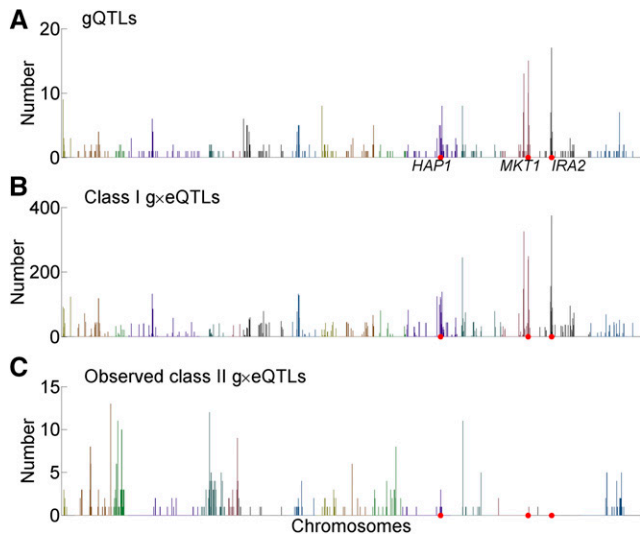


Figure 2 Genomic distributions of (A) gQTLs, (B) class I g×eQTLs, and (C) observed class II g×eQTLs. The genome is divided into 7500-nucleotide bins. The total number of gQTLs from all 47 environments, the total number of class I g×eQTLs from all 1081 pairs of environments, and the total number of observed class II g×eQTLs from all 1081 pairs of environments are plotted for each bin. The 16 chromosomes are colored differently. Three genes referred to in the main text are marked according to their genomic locations.

G×E involving natural genetic polymorphisms (Fry *et al.* 1998; El-Soda *et al.* 2014; Dillon *et al.* 2016), but the numbers of cases were all small, and thus the generality of these observations is unclear. The large collection of yeast data analyzed here appears to show the same pattern. A g×eQTL is considered antagonistic between two environments if the BY allele is fitter than the RM allele in one environment, while the RM allele is fitter than the BY allele in the other environment, even if the difference is statistically significant in neither environment. Otherwise, the g×eQTL is considered concordant between the two environments. Thus, purely by chance, we would expect a g×eQTL to be equally likely to be antagonistic and concordant. However, on average only 28% of class I g×eQTLs are antagonistic, significantly lower than the null expectation ($P < 10^{-250}$, binomial test; Figure 4A). Among the observed class II g×eQTLs, 94% are antagonistic, which is not unexpected, because a concordant g×eQTL should have a significant effect in at least one of the environments, and thus is unlikely to be of class II. Because class I g×eQTLs substantially outnumber class II g×eQTLs (Figure 2), only 37% of all g×eQTLs are antagonistic ($P < 10^{-171}$, binomial test), under the assumption that antagonism is equally frequent among the observed and extrapolated class II g×eQTLs.

Large-effect QTLs are more likely than small-effect QTLs to be antagonistic

A previous study of yeast gene deletions identified many antagonisms between environments (Qian *et al.* 2012), seemingly contrasting the scarcity of antagonism of natural polymorphisms surveyed in the present study. Because gene deletions should, on average, have larger phenotypic effects than natural polymorphisms, a potential explanation of the

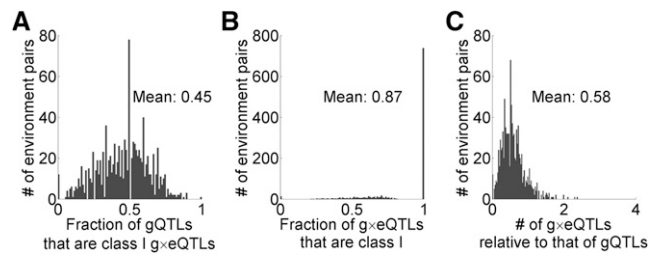


Figure 3 Relative numbers of g×eQTLs and gQTLs from all pairs of environments. (A) Frequency distribution of the fraction of unique gQTLs identified from two individual environments that are class I g×eQTLs for the pair of environments. (B) Frequency distribution of the fraction of all g×eQTLs (*i.e.*, class I + extrapolated class II) that are class I. (C) Frequency distribution of the ratio between the number of all g×eQTLs for a pair of environments and the total number of unique gQTLs identified in the two environments.

disparity in the frequency of antagonism may be that large-effect mutations are more likely than small-effect mutations to be antagonistic. To directly test this hypothesis, for each gQTL, we counted the number of environments where its effect is opposite to the effect in the environment where the gQTL was detected. Indeed, the larger the effect of a gQTL, the higher the likelihood that it has an antagonistic effect in another environment ($\rho = 0.14$, $P < 10^{-4}$; Figure 4B).

Prevalence of antagonism varies among environments

To study whether antagonism is enriched in certain environments, for each pair of environments, we calculated the fraction of class I g×eQTLs that are antagonistic. If this fraction is 0, we say that this pair of environments is nonantagonistic to each other. Similarly, if this fraction ≥ 0.5 , these two environments are highly antagonistic to each other. We counted the number of times that each environment is said to be nonantagonistic, and the number of times that it is said to be highly antagonistic to another environment. We then computed the mean number of times that an environment is nonantagonistic, and the mean number of times that an environment is highly antagonistic. Environments showing ≥ 2 times the mean number of nonantagonism are galactose, caffeine, 4-hydroxybenzaldehyde, calcium chloride, mannose, menadione, and YNB (Figure 4C), whereas those exhibiting ≥ 2 times the mean number of high antagonism are cadmium chloride, copper, hydrogen peroxide, and cycloheximide (Figure 4D). A potential explanation of the among-environment variation in the prevalence of antagonism is that antagonisms may have been resolved by natural selection in commonly encountered environments, but not in rarely encountered environments (Qian *et al.* 2012). However, to what extent the environments in Figure 4C are more common than the environments in Figure 4D is unknown, due to the paucity of ecological information for yeast. Another possibility, not mutually exclusive with the above, is that some environments are more dissimilar to other environments, and hence exhibit more antagonism. In support of the latter hypothesis, the fraction of antagonistic class I g×eQTLs between two environments negatively correlates with their environment similarity ($\rho = -0.61$, $P < 10^{-110}$).

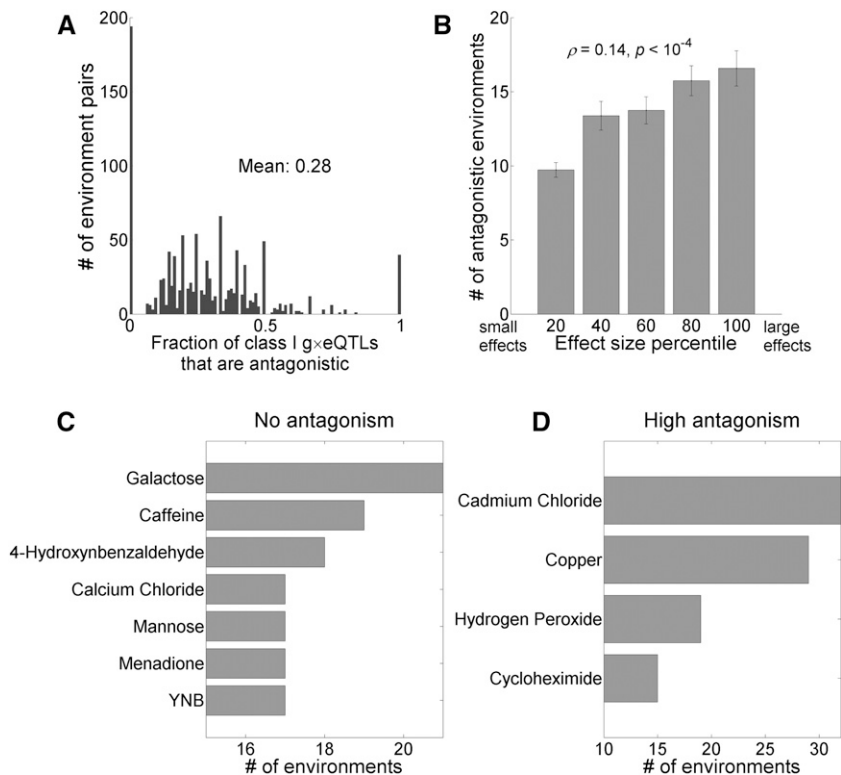


Figure 4 Patterns of antagonistic G×E. (A) Frequency distribution of the fraction of class I g×eQTLs that are antagonistic. (B) gQTLs with large effects in the environments where they are identified are more likely than small-effect gQTLs to have antagonistic effects in another environment. Error bars indicate one SE. The rank correlation ρ and associated P -value are based on the unbinned data. (C) Environments that are under-represented with antagonistic g×eQTLs with other environments. The X -axis shows the number of environments with which an environment listed on the Y -axis has no antagonistic class I g×eQTL. (D) Environments that are enriched with antagonistic g×eQTLs with other environments. The X -axis shows the number of environments with which an environment listed on the Y -axis has >50% of class I g×eQTLs being antagonistic.

Distributions of gQTLs and g×eQTLs across the genome

To understand the molecular basis of G×E, we first categorized all 28,220 SNPs between BY and RM strains into coding SNPs, intronic SNPs, and intergenic SNPs. We merged gQTLs from all environments, and merged class I g×eQTLs from all environment pairs. A gQTL or g×eQTL was counted as many times as it appears in the merged list. Table 1 summarizes the results of enrichment tests for each genomic category. Compared with all SNPs, gQTLs are not significantly different in frequency distribution among coding, intronic, and intergenic regions (Table 1). Relative to gQTLs, class I g×eQTLs are twofold more likely to be in introns ($P = 2.2 \times 10^{-7}$; Table 1), suggesting that yeast introns are more important in regulating environment-dependent growth rates than environment-independent growth rates.

We also analyzed the distributions of gQTLs and g×eQTLs among synonymous, nonsynonymous, and nonsense SNPs within coding regions. A synonymous SNP does not alter the amino acid encoded by the codon where the SNP resides, whereas a nonsynonymous SNP alters the amino acid. A nonsense SNP changes a sense codon in one strain to a stop codon in another. Relative to all SNPs, gQTLs are more likely to occur at nonsynonymous SNPs (1.125-fold, $P = 0.03$), and are less likely to occur at synonymous SNPs (0.896-fold, $P = 0.02$). This observation is not unexpected, because nonsynonymous mutations are more likely than synonymous mutations to have phenotypic effects. Relative to gQTLs, g×eQTLs are more likely to occur at synonymous SNPs (1.070-fold, $P = 9.6 \times 10^{-9}$), but are less likely to occur at nonsynonymous (0.935-fold, $P = 2.5 \times 10^{-7}$) and non-

sense (0.826-fold, $P = 0.0265$) SNPs, suggesting that non-synonymous and nonsense mutations tend to have universal rather than environment-specific growth effects, when compared with synonymous mutations. Among all g×eQTLs, we analyzed only class I g×eQTLs here, because the number of class II g×eQTLs is small, and because our simulation (Table S1) showed that mapping is less precise for class II g×eQTLs.

Note that, because the gQTLs and g×eQTLs identified may not be causal SNPs but are simply linked with causal SNPs, the above analysis has a lower statistical power than when causal SNPs are used in the analysis. In our simulation, >31% of gQTLs and >29% of class I g×eQTLs are mapped to causal SNPs (Table S1), suggesting that a sizable proportion of mapped sites are causal, explaining why our test is not entirely powerless. Thus, the significant results obtained are likely to be genuine, and the conclusions conservative.

Different GO distributions of gQTLs and g×eQTLs

GO annotation is organized into three domains: cellular component, molecular function, and biological process (Ashburner *et al.* 2000). Each domain contains many GO terms, which may be a word or string of words related to gene function. A gene is annotated for all three domains and one to many terms in each domain, on the basis of its product and function. We examined the enrichment of gQTLs and g×eQTLs for GO domains and terms (Table 2). Note that intergenic SNPs were assigned to their closest genes. We compared gQTLs to the background of all SNPs, and compared class I g×eQTLs to the background of all gQTLs, using binomial tests followed by Bonferroni corrections with a

Table 1 Distributions of gQTLs and class I g×eQTLs across various genomic regions

Genomic Regions	All SNPs ^a	All gQTLs ^b		Class I g×eQTLs ^c	
	Frequency	Frequency	P-value ^d	Frequency	P-Value ^e
Intronic	0.008	0.004	0.2804	0.008	2.2×10^{-7}
Intergenic	0.331	0.344	0.2548	0.324	1.0×10^{-5}
Coding	0.656	0.643	0.2927	0.665	2.3×10^{-6}
Synonymous	0.558	0.500	0.0234	0.535	9.6×10^{-9}
Nonsynonymous	0.425	0.478	0.0265	0.447	2.5×10^{-7}
Nonsense	0.018	0.023	0.1859	0.019	0.0265

^a Total number of SNPs is 28,220.

^b Total number of gQTLs is 482.

^c Total number of class I g×eQTLs is 9952.

^d Comparison with all SNPs using a binomial test.

^e Comparison with all gQTLs using a binomial test.

corrected $P = 0.05$ as the cutoff. Compared with all SNPs, gQTLs are not enriched in any GO domain, but are significantly enriched in 24 GO terms (File S3). gQTLs are not underrepresented in any GO domain or GO term. These results suggest that gQTLs are overall annotated with more functions than average SNPs. Relative to gQTLs, class I g×eQTLs are enriched in the GO domain cellular component ($P = 0.028$), suggesting that proteins encoded by g×eQTLs have relatively more locations in the cell, or are relatively better annotated for cellular component. Class I g×eQTLs are significantly underrepresented in biological process ($P = 4.2 \times 10^{-6}$) and molecular function ($P = 3 \times 10^{-6}$), when compared with gQTLs. Strikingly, of the 848 GO terms that contain at least one gQTL, g×eQTLs are enriched in 137 of them, and are underrepresented in 139 (File S3). Of the GO terms enriched in gQTLs, four terms are further enriched in g×eQTLs (Table 2), and four are underrepresented. Thus, the functional distributions of gQTLs and class I g×eQTLs are quite different, despite the fact that the latter constitutes a large subset of the former. One potential bias in the above GO enrichment analysis of gQTLs is that SNPs are not distributed evenly along genes and chromosomes. To rectify this problem, we also tested GO enrichment of gQTLs against all genes, instead of all SNPs, by assigning each gQTL to its closest gene. The enriched GO terms (File S4), however, remained largely the same.

Antagonistic and concordant g×eQTLs have different genomic and functional enrichments

Comparing antagonistic and concordant class I g×eQTLs, we found no significant difference in their frequency distributions among coding, intronic, and intergenic regions (Table 3). However, within coding regions, antagonistic g×eQTLs are enriched at synonymous ($P = 9.7 \times 10^{-9}$, chi-squared test) and nonsense SNPs ($P = 2.7 \times 10^{-7}$), but underrepresented at nonsynonymous SNPs ($P = 7.6 \times 10^{-13}$), when compared with concordant g×eQTLs.

Antagonistic and concordant g×eQTLs show significantly different enrichments for two GO domains, biological process (adjusted $P = 4.2 \times 10^{-4}$, chi-squared test; File S5), and cellular component (adjusted $P = 1.4 \times 10^{-6}$). They are

also significantly different in 187 of 907 GO terms that have at least one occurrence in class I g×eQTLs (File S5). Interestingly, two (ribosomal small subunit biogenesis and 90S preribosome) of the five GO terms significantly enriched in both gQTLs and g×eQTLs are the top two terms that differ significantly between antagonistic and concordant g×eQTLs; they each occur 325 times in concordant g×eQTLs but 0 times in antagonistic g×eQTLs. This result suggests that, although differences in translation underlie g×eQTLs, these differences mostly have concordant G×E effects.

Ignoring G×E causes missing heritability

“Missing heritability” refers to the gap between the phenotypic variance explained by GWAS results and those estimated from classical heritability methods (Zaitlen and Kraft 2012), and is a prominent problem in the study of human complex traits that has attracted much attention (Manolio *et al.* 2009; Eichler *et al.* 2010). G×E has been proposed as a potential cause for the missing heritability problem (Manolio *et al.* 2009; Eichler *et al.* 2010). Because heritability is classically estimated from relatives such as by comparing monozygotic (MZ) and dizygotic (DZ) twins, the effect of environmental heterogeneity for a twin is canceled in the comparison between MZ and DZ twins, and has no effect on the heritability estimate. However, in human GWAS, the environmental effect and G×E effect are rarely controlled, which could lower the power in identifying the underlying genetic variants, and render the estimation of effect size inaccurate. To quantitatively evaluate the contribution of ignoring G×E to the missing heritability problem, we conducted a simulation using the yeast data. That is, for one half of the segregants, we used their phenotypes measured in one environment, but for the other half of the segregants, we used their phenotypes measured in another environment. We then attempted to identify gQTLs as if all segregants were phenotyped in the same environment. We did this simulation for 100 random pairs of environments. An example is provided in Figure 5A, where the phenotype data are from YNB at 30° and YPD at 37°. Ten and eight gQTLs were identified from 1005 segregants in YNB and YPD, respectively, but only two gQTLs were identified from the

Table 2 Significantly overrepresented gene ontology (GO) domains and terms

GO Category	All SNPs ^a	All gQTLs ^b		Class I g×eQTLs ^c	
	Frequency	Frequency	P-value ^d	Frequency	P-Value ^d
GO domains					
Cellular component	0.692	0.685	1	0.696	0.028
Biological process	0.755	0.790	0.087	0.771	4.2×10^{-6}
Molecular function	0.819	0.842	0.23	0.825	3.0×10^{-6}
GO terms ^e					
GDP binding	7.1×10^{-5}	6.2×10^{-3}	1.9×10^{-4}	1.2×10^{-2}	3.2×10^{-9}
Sequence-specific DNA binding transcription factor activity	2.5×10^{-2}	6.4×10^{-2}	4.1×10^{-3}	7.5×10^{-2}	1.3×10^{-4}
Ribosomal small subunit biogenesis	3.4×10^{-3}	2.7×10^{-2}	7.7×10^{-6}	3.3×10^{-2}	7.3×10^{-3}
90S preribosome	5.3×10^{-3}	2.7×10^{-2}	1.6×10^{-3}	3.3×10^{-2}	7.3×10^{-3}

^a Total number of SNPs is 28,220.

^b Total number of gQTLs is 482.

^c Total number of class I g×eQTLs is 9952.

^d Based on a binomial test followed by multiple-testing correction. gQTLs are compared with all SNPs while class I g×eQTLs are compared with all gQTLs.

^e Shown are GO terms significantly enriched in both gQTLs (relative to all SNPs) and class I g×eQTLs (relative to gQTLs).

mixture of the phenotype data of 502 segregants in YNB and 503 segregants in YPD, although these two gQTLs are a subset of the 18 gQTLs identified from the individual environments. When the phenotype data of the 1005 segregants are all from either YNB or YPD but not both, the identified gQTLs together can explain, on average, 54% of the total phenotypic variance observed among the segregants. This number reduces to 26% when the mixed phenotype data are used (green dots in Figure 5A). To distinguish between the environmental effect and G×E effect on gQTL identification, we conducted another analysis, in which the phenotypic value of a segregant in an environment is defined by the difference between its raw phenotypic value and the mean phenotypic value of all segregants in that environment. We then mixed these normalized phenotypic values from two environments to identify gQTLs. We found that such normalization improves gQTL identification, because the number of gQTLs identified rises to six, although this number is still smaller than when homogenous data are used. The total variance of normalized phenotypes explained rises to 42%. The remaining difference between this result (light salmon symbols in Figure 5A) and the original result (blue and red symbols in Figure 5A) is attributable to G×E.

On average, across the 100 random pairs of environments, the identified gQTLs explain 40% of the total phenotypic variance among segregants under one environment. When mixed phenotypic data from two environments are used, this number drops to 10% (Figure 5B). When phenotypic data are normalized by the mean phenotypic value of the environment, the fraction of phenotypic variance explained is 23% (Figure 5B). Hence, in this dataset, environmental effects and G×E effects have similar amounts of contribution to missing heritability. We also conducted 100 simulations where the phenotype data are generated from 5 and 10 environments, respectively. As the number of environments increases, the amount of missing heritability rises, the contribution of G×E to missing heritability increases, and the contribution of environmental effects decreases (Figure 5B).

We further calculated the distances between the gQTLs identified using the mixed phenotypes from two environments, and the nearest gQTLs identified using phenotypes from individual environments for all 100 random pairs of environments (Figure 5C). We found that, although noise is larger in mixed environments, the identified sites are generally closely linked to the gQTLs identified from individual environments. This is true both with and without controlling the environmental effect. What types of gQTLs are underdetected using mixed phenotype data? On the basis of the same 100 pairs of environments examined, we found that, on average, 23.6% of gQTLs having the same direction of effect in the two environments, and 12.7% of gQTLs having opposite directions of effect were detected using the mixed data, when the environmental effect is uncontrolled ($P = 7.1 \times 10^{-14}$, *t*-test of equal probability of detection for the two groups of gQTLs). These numbers increase to 52.0 and 33.8%, respectively, upon the control of the environmental effect ($P < 8.5 \times 10^{-6}$). Thus, while all gQTLs are underdetected using mixed phenotype data, those with opposite effects in the two environments suffer more than those with the same direction of effect.

In human GWAS, larger and larger samples are being used, despite the fact that enlarging samples likely increases the environmental heterogeneity of the sample. To study this effect, we merged the phenotype data from all 47 environments, resulting in a sample of 42,781 individuals; this number is $< 47 \times 1005 = 47,325$ because not all 1005 individuals had growth data in all 47 environments. Using this very large sample, we were able to map 21 gQTLs—more than the number of gQTLs mapped individually in 46 of the 47 environments. Some of the mapped gQTLs overlapped with the gQTLs frequently identified in individual environments (Figure S1), suggesting that using large samples in GWAS might help identify influential loci that have effects in multiple environments. Nevertheless, the fraction of phenotypic variance explained by all mapped sites is only 2.5%, similar to that when a sample of 1005 segregants, each fifth originating from a different environment, is used, and much lower than that when a sample of

Table 3 Distributions of antagonistic and concordant class $g \times e$ QTLs across various genomic regions

Genomic Regions	Antagonistic		Concordant		P-Value ^a
	Frequency	Occurrences	Frequency	Occurrences	
Intronic	0.0059	16	0.0084	61	0.2083
Intergenic	0.3370	910	0.3194	2316	0.0939
Coding	0.6556	1770	0.6685	4848	0.2236
Synonymous	0.5927	1049	0.5132	2488	9.7×10^{-9}
Nonsynonymous	0.3740	662	0.4730	2293	7.6×10^{-13}
Nonsense	0.0333	59	0.0138	67	2.7×10^{-7}

^a Based on a chi-squared test.

1005 segregants from the same environment is used (Figure 5B). Clearly, the missing heritability problem worsens when enlarging samples also increases environmental heterogeneity.

Discussion

We conducted a systematic analysis of interaction between natural genetic variants and environments in yeast growth, and identified numerous $g \times e$ QTLs. The average number of $g \times e$ QTLs identified between two environments is 0.58 times the number of unique gQTLs identified in the two environments, indicating a high abundance of $G \times E$. It is debated whether testing all SNPs, or testing only those with effects in at least one of the environments concerned, is more suitable for $G \times E$ detection (Duncan and Keller 2011; Uher 2014). Our computer simulation showed that using the latter approach has the benefit of lowering the false discovery rate and increasing the chance of finding causal variants. Although our simulation also indicated that the latter approach has a higher false negative rate than the former approach, our yeast data analysis found that 88% of $g \times e$ QTLs could be identified from gQTLs. Similar results were obtained when the larger dataset of Bloom *et al.* (2015) was analyzed (Figure S2). Together, these findings support the current practice in human genetics of using genes or QTLs known to have effects in at least one of the environments concerned as candidates in the study of $G \times E$. The gQTL mapping method and $G \times E$ detection method developed here are expected to suit other similar large-scale studies of $G \times E$. In our computer simulation, we found that both Storey and Tibshirani Q-value (Storey and Tibshirani 2003) and permutation Q-value (Doerge and Churchill 1996) underestimate the false discovery rate. This underestimation may be a general problem in linkage mapping of complex traits, suggesting the importance of using computer simulation to assess false discovery rates.

We found that most $G \times E$ interactions are concordant, suggesting that the fitness landscapes in different environments examined are positively correlated, such that a mutation that is beneficial in one tested environment tends to be beneficial in other tested environments. Nevertheless, we detected a few environments with unusually high degrees of antagonistic $G \times E$, such as those with trace minerals or heavy metals. Because we observed a negative correlation between

the fraction of antagonistic $g \times e$ QTLs and environmental similarity, it is likely that these antagonism-rich environments are relatively dissimilar to the other environments examined. The antagonism-rich environments may also be rarely encountered by yeast in nature such that antagonism has not had chance to be resolved by natural selection. We did not attempt to verify the disparity in antagonism among environments using the data of Bloom *et al.* (2015), because only 3 of the 11 environments in Figure 4, C and D are included in this dataset. The fact that the extent of antagonism depends on the tested environments illustrates the importance in carefully choosing environments in testing the potential antagonism of beneficial mutations observed in experimental evolution (Ostrowski *et al.* 2005; Wenger *et al.* 2011; Bedhomme *et al.* 2012; Dillon *et al.* 2016).

We observed that large-effect gQTLs identified in one environment are more likely than small-effect gQTLs to have antagonistic effects in another environment, reminiscent of the common belief and a prediction of Fisher's geometric model (Fisher 1930) that large-effect mutations are more likely than small-effect mutations to be deleterious. Our observation predicts that the prevalence of detected antagonism will decrease with the power of $g \times e$ QTL mapping, because, as the power increases, $g \times e$ QTLs of smaller and smaller effects are mapped. This prediction is confirmed by using the larger dataset from Bloom *et al.* (2015), where the fraction of antagonistic $g \times e$ QTLs was found to be even lower (Figure S3). Note, however, that we studied growth rate, a primary component of fitness, in this work. For traits that are either unrelated to, or only loosely correlated with, fitness, antagonism patterns may be different because they are not subject to the same degree of natural selection.

We tested the enrichment of different functional sites of the yeast genome as well as different GO categories in $g \times e$ QTLs and gQTLs. We found that gQTLs are enriched with nonsynonymous SNPs, similar to the collective finding from human GWAS studies (Hindorff *et al.* 2009). Relative to gQTLs, $g \times e$ QTLs are more likely to occur at intronic SNPs. We confirmed the enrichment of nonsynonymous SNPs in gQTLs and enrichment of intronic SNPs in $g \times e$ QTLs (Table S3) using the data from Bloom *et al.* (2015). Concordant and antagonistic $g \times e$ QTLs also have different distributions among the three categories of coding SNPs, with concordant $g \times e$ QTLs enriched at nonsynonymous SNPs, and antagonistic $g \times e$ QTLs enriched

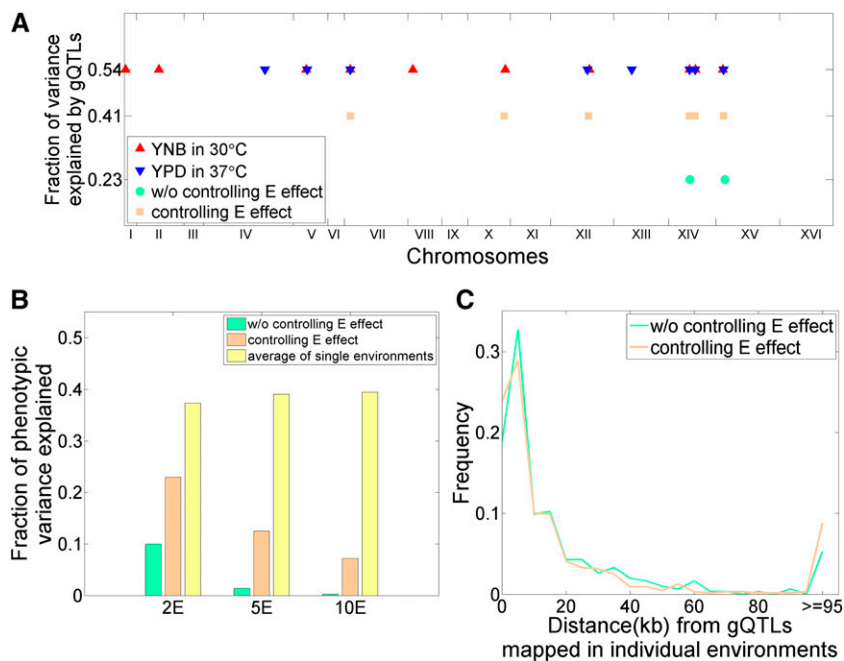


Figure 5 Ignoring G×E causes missing heritability. (A) The genomic distributions of gQTLs identified from phenotypes measured in one environment and those measured in two environments (50% segregants from each environment), respectively. Y-axis shows the fraction of phenotypic variance explained by the identified gQTLs under each mapping scheme. E effect, environmental effect. Without controlling E effect means that neither environmental effect nor G×E is considered in mapping. Controlling E effect means environmental effect but not G×E is considered in mapping. (B) Average fraction of phenotypic variance explained by gQTLs (r^2) decreases as the phenotypic data used originate from more environments. The average narrow-sense heritability is 0.55. 2E, phenotypic data are from a mixture of two environments; 5E, phenotypic data are from a mixture of five environments; 10E, phenotypic data are from a mixture of 10 environments. Results are summarized from 100 random sets of 2, 5, and 10 environments, respectively. (C) Frequency distribution of the distance between gQTLs identified using mixed phenotypes from two environments and those identified using phenotypes from individual environments. The results are summarized from 100 random pairs of environments.

at synonymous and nonsense SNPs. Data from Bloom *et al.* (2015) showed the same patterns, except that the distribution of nonsense SNPs were not significantly different between concordant and antagonistic g×eQTLs (Table S4). These results suggest a different molecular basis of concordant and antagonistic G×E. We also found g×eQTLs to be enriched in GO terms on ribosome and translation (Table 2), which is potentially related to the aforementioned enrichment in introns, because introns are concentrated in ribosomal protein genes in yeast (Parenteau *et al.* 2011). The correlation between ribosomal protein gene expression and growth rate is well known (Mager and Planta 1991), and the comparisons between gQTLs and g×eQTLs and between antagonistic and concordant g×eQTLs using the data from Bloom *et al.* (2015) suggest the possibility that intronic SNPs affect ribosomal protein gene expression, which potentially affects growth rate differently in different environments. Specifically, introns from four genes (*TUB3*, *PFY1*, *RPL34B*, and *RPL40B*) are found to harbor gQTLs. While concordant intronic g×eQTLs are found in all of the four genes, antagonistic intronic g×eQTLs are found only in the two ribosomal protein genes (*RPL34B* and *RPL40B*). Using the data of Bloom *et al.* (2015), we found that 38 GO terms are enriched in gQTLs, while only one GO term is underrepresented, confirming that gQTLs are overall annotated with more functions than average SNPs.

Our yeast data-based simulation of mixed environments revealed the importance of considering G×E in QTL mapping, and, by extension, association studies. Neglecting environmental heterogeneity in the data substantially reduces the number of QTLs identified and results in missing heritability. Many human genetic association studies ignore the fact that different individuals have different environments, and our results suggest that failure to account for environmental het-

erogeneity could be a primary reason underlying the missing heritability phenomenon. Another commonly cited cause of missing heritability is epistasis, or gene-by-gene interaction (G×G). But recent studies found that failure to consider G×G is not a primary cause of missing heritability (Bloom *et al.* 2013, 2015). In model organism studies, where the environment tends to be well controlled, missing heritability tends to be mild. But, in human GWAS, where environments are hard to control, missing heritability is severe (Eichler *et al.* 2010). This contrast, coupled with our simulation results, suggests that missing heritability in human GWAS may be due primarily to ignoring environmental factors and/or G×E. We showed in our simulation that using very large samples could help identify more influential loci when compared with small samples of environmental homogeneity, but the “missing heritability” problem is exacerbated if enlarging a sample means increasing the environmental heterogeneity of the sample. Although it is impossible to have different human individuals living in exactly the same environment, even partially controlling environments helps identify disease-associated alleles. For example, in GWAS of type II diabetes, controlling for obesity in statistical analysis helps identify new disease-associated variants (Zeggini *et al.* 2008). This kind of controlling of environmental/physiological factors will help identify new trait-associated genetic variants and reduce missing heritability. Notwithstanding, because classical estimation of heritability is minimally affected by environmental heterogeneity, while modern GWAS is subject to potentially high environmental heterogeneity, the “missing heritability” due to this difference may be considered fictional (Heckerman *et al.* 2016). Better estimation of heritability by considering environmental heterogeneity will help gauge the true missing heritability in GWAS (Heckerman *et al.* 2016).

Acknowledgments

We thank the Kruglyak laboratory for sharing the yeast segregant genotype and phenotype data, and Soochin Cho, Wei-Chin Ho, Chuan Li, Jian-Rong Yang, and two anonymous reviewers for valuable comments. This work was supported by the United States National Institutes of Health research grant R01GM103232 to J.Z.

Literature Cited

- Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler *et al.*, 2000 Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: 25–29.
- Bedhomme, S., G. Lafforgue, and S. F. Elena, 2012 Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. *Mol. Biol. Evol.* 29: 1481–1492.
- Bloom, J. S., I. M. Ehrenreich, W. T. Loo, T. L. Lite, and L. Kruglyak, 2013 Finding the sources of missing heritability in a yeast cross. *Nature* 494: 234–237.
- Bloom, J. S., I. Kottenko, M. J. Sadhu, S. Treusch, F. W. Albert *et al.*, 2015 Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. *Nat. Commun.* 6: 8712.
- Brown, J. A., G. Sherlock, C. L. Myers, N. M. Burrows, C. Deng *et al.*, 2006 Global analysis of gene function in yeast by quantitative phenotypic profiling. *Mol Syst Biol* 2: 2006.0001.
- Byrd, A. L., and S. B. Manuck, 2014 MAOA, childhood maltreatment, and antisocial behavior: meta-analysis of a gene-environment interaction. *Biol. Psychiatry* 75: 9–17.
- Caspi, A., J. McClay, T. E. Moffitt, J. Mill, J. Martin *et al.*, 2002 Role of genotype in the cycle of violence in maltreated children. *Science* 297: 851–854.
- Caspi, A., K. Sugden, T. E. Moffitt, A. Taylor, I. W. Craig *et al.*, 2003 Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301: 386–389.
- Caspi, A., T. E. Moffitt, M. Cannon, J. McClay, R. Murray *et al.*, 2005 Moderation of the effect of adolescent-onset cannabis use on adult psychosis by a functional polymorphism in the catechol-O-methyltransferase gene: longitudinal evidence of a gene X environment interaction. *Biol. Psychiatry* 57: 1117–1127.
- Chamaillard, M., D. Philpott, S. E. Girardin, H. Zouali, S. Lesage *et al.*, 2003 Gene-environment interaction modulated by allelic heterogeneity in inflammatory diseases. *Proc. Natl. Acad. Sci. USA* 100: 3455–3460.
- Cubillos, F. A., E. Billi, E. Zörgö, L. Parts, P. Fargier *et al.*, 2011 Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol. Ecol.* 20: 1401–1413.
- Dillon, M. M., N. P. Rouillard, B. Van Dam, R. Gallet, and V. S. Cooper, 2016 Diverse phenotypic and genetic responses to short-term selection in evolving *Escherichia coli* populations. *Evolution* 70: 586–599.
- Doerge, R. W., and G. A. Churchill, 1996 Permutation tests for multiple loci affecting a quantitative character. *Genetics* 142: 285–294.
- Dudley, A. M., D. M. Janse, A. Tanay, R. Shamir, and G. M. Church, 2005 A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol Syst Biol* 1: 2005.0001.
- Duncan, L. E., and M. C. Keller, 2011 A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am. J. Psychiatry* 168: 1041–1049.
- Ehrenreich, I. M., J. Bloom, N. Torabi, X. Wang, Y. Jia *et al.*, 2012 Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet.* 8: e1002570.
- Eichler, E. E., J. Flint, G. Gibson, A. Kong, S. M. Leal *et al.*, 2010 Missing heritability and strategies for finding the underlying causes of complex disease. *Nat. Rev. Genet.* 11: 446–450.
- El-Soda, M., M. Malosetti, B. J. Zwaan, M. Koornneef, and M. G. Aarts, 2014 Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends Plant Sci.* 19: 390–398.
- Fisher, R. A., 1930 *The Genetic Theory of Natural Selection*. Clarendon, Oxford.
- Flint, J., and T. F. Mackay, 2009 Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* 19: 723–733.
- Fry, J. D., S. V. Nuzhdin, E. G. Pasyukova, and T. F. Mackay, 1998 QTL mapping of genotype-environment interaction for fitness in *Drosophila melanogaster*. *Genet. Res.* 71: 133–141.
- Gagneur, J., O. Stegle, C. Zhu, P. Jakob, M. M. Tekkedil *et al.*, 2013 Genotype-environment interactions reveal causal pathways that mediate genetic effects on phenotype. *PLoS Genet.* 9: e1003803.
- Gerke, J., K. Lorenz, S. Ramnarine, and B. Cohen, 2010 Gene-environment interactions at nucleotide resolution. *PLoS Genet.* 6: e1001144.
- Heckerman, D., D. Gurdasani, C. Kadie, C. Pomilla, T. Carstensen *et al.*, 2016 Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci. USA* 113: 7377–7382.
- Hillenmeyer, M. E., E. Fung, J. Wildenhain, S. E. Pierce, S. Hoon *et al.*, 2008 The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* 320: 362–365.
- Hindorf, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta *et al.*, 2009 Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* 106: 9362–9367.
- Hood, L., J. R. Heath, M. E. Phelps, and B. Lin, 2004 Systems biology and new technologies enable predictive and preventative medicine. *Science* 306: 640–643.
- Hunter, D. J., 2005 Gene-environment interactions in human diseases. *Nat. Rev. Genet.* 6: 287–298.
- Kendler, K. S., K. Sundquist, H. Ohlsson, K. Palmér, H. Maes *et al.*, 2012 Genetic and familial environmental influences on the risk for drug abuse: a national Swedish adoption study. *Arch. Gen. Psychiatry* 69: 690–697.
- Li, Y., O. A. Alvarez, E. W. Gutteling, M. Tijsterman, J. Fu *et al.*, 2006 Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* 2: e222.
- Luck, T., S. Riedel-Heller, M. Lupp, B. Wiese, M. Köhler *et al.*, 2014 Apolipoprotein E epsilon 4 genotype and a physically active lifestyle in late life: analysis of gene–environment interaction for the risk of dementia and Alzheimer’s disease dementia. *Psychol. Med.* 44: 1319–1329.
- Luo, G., J. Kim, and K. Song, 2014 The C-terminal domains of human neurofibromin and its budding yeast homologs Ira1 and Ira2 regulate the metaphase to anaphase transition. *Cell Cycle* 13: 2780–2789.
- Mager, W. H., and R. J. Planta, 1991 Coordinate expression of ribosomal protein genes in yeast as a function of cellular growth rate, pp. 181–187 in *Molecular Mechanisms of Cellular Growth*. Springer, New York.
- Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454: 479–485.
- Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorf *et al.*, 2009 Finding the missing heritability of complex diseases. *Nature* 461: 747–753.

- Matsui, T., and I. M. Ehrenreich, 2016 Gene-environment interactions in stress response contribute additively to a genotype-environment interaction. *PLoS Genet.* 12: e1006158.
- Moffitt, T. E., A. Caspi, and M. Rutter, 2005 Strategy for investigating interactions between measured genes and measured environments. *Arch. Gen. Psychiatry* 62: 473–481.
- Ostrowski, E. A., D. E. Rozen, and R. E. Lenski, 2005 Pleiotropic effects of beneficial mutations in *Escherichia coli*. *Evolution* 59: 2343–2352.
- Ottman, R., 1996 Gene-environment interaction: definitions and study designs. *Prev. Med.* 25: 764–770.
- Padyukov, L., C. Silva, P. Stolt, L. Alfredsson, and L. Klareskog, 2004 A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum.* 50: 3085–3092.
- Parenteau, J., M. Durand, G. Morin, J. Gagnon, J.-F. Lucier *et al.*, 2011 Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* 147: 320–331.
- Qian, W., D. Ma, C. Xiao, Z. Wang, and J. Zhang, 2012 The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Reports* 2: 1399–1410.
- Risch, N., R. Herrell, T. Lehner, K.-Y. Liang, L. Eaves *et al.*, 2009 Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA* 301: 2462–2471.
- Smith, E. N., and L. Kruglyak, 2008 Gene-environment interaction in yeast gene expression. *PLoS Biol.* 6: e83.
- Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* 100: 9440–9445.
- Thorgeirsson, T. E., F. Geller, P. Sulem, T. Rafnar, A. Wiste *et al.*, 2008 A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 452: 638–642.
- Uher, R., 2014 Gene-environment interactions in common mental disorders: an update and strategy for a genome-wide search. *Soc. Psychiatry Psychiatr. Epidemiol.* 49: 3–14.
- Ungerer, M. C., S. S. Halldorsdottir, M. D. Purugganan, and T. F. Mackay, 2003 Genotype-environment interactions at quantitative trait loci affecting inflorescence development in *Arabidopsis thaliana*. *Genetics* 165: 353–365.
- Wenger, J. W., J. Piotrowski, S. Nagarajan, K. Chiotti, G. Sherlock *et al.*, 2011 Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLoS Genet.* 7: e1002202.
- Wilkening, S., G. Lin, E. S. Fritsch, M. M. Tekkedil, S. Anders *et al.*, 2014 An evaluation of high-throughput approaches to QTL mapping in *Saccharomyces cerevisiae*. *Genetics* 196: 853–865.
- Yadav, A., K. Dhole, and H. Sinha, 2016 Genetic regulation of phenotypic plasticity and canalisation in yeast growth. *PLoS One* 11: e0162326.
- Zaitlen, N., and P. Kraft, 2012 Heritability in the genome-wide association era. *Hum. Genet.* 131: 1655–1664.
- Zeggini, E., L. J. Scott, R. Saxena, B. F. Voight, J. L. Marchini *et al.*, 2008 Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* 40: 638–645.

Communicating editor: B. A. Payseur