Supplementary Materials:

Supplementary Materials and Methods

Populations sampled

The dataset used to estimate genetic variance for gene expression in *S. cerevisiae* included seven isolates sampled from Italian vineyards in 1993/1994, two isolates sampled from oak trees in Pennsylania in 1999, and one classic lab strain S288c(Fay et al. 2004). Genetic relatedness of these strains assessed in reference (Fay and Benavides 2005) suggests that S288c and the oak isolates were more distantly related to the wine isolates than the wine isolates were to one another so V_s was estimated from both the wine isolates alone and all 10 strains to assess whether reduced recombination between less closely related strains affected estimates of selection. Yeast isolates were cultured in rich YPD media and RNA was extracted after exponential growth of clonal cultures in which the OD_{600} was between 0.8 and 1.

The dataset used to estimated genetic variance for gene expression in *D. melanogaster* included adult males from sixteen strains of *D. melanogaster* from natural populations in the Netherlands and Zimbabwe (Hutter et al. 2008). The flies were reared on cornmeal-molasses medium at 22°C under a 15/9 h light/dark cycle. Whole body RNA was extracted from 70-75 flies of 4-6 days old.

The dataset used to estimate genetic variance for gene expression in *C. elegans* included 5 cosmopolitan *C. elegans* natural isolate lines. Worms were cultured on RNGM agarose plates under standard protocols and RNA was extracted from developmentally synchronized worms at the young adult stage (Denver et al. 2005).

Model averaging procedure

To incorporate uncertainty about the correct evolutionary model into our assessment of the phenotype space of gene expression, we performed model averaging (Hoeting et al. 1999) to estimate V_s for evaluation of the inequality $20\,\mu V_s < \alpha^2$. For each gene, the proportion of parameter samples from their uncertainty distributions that satisfied this inequality for the Gaussian and House of Cards models was calculated to generate a probability for the fit of each model. Starting from the Maximum Entropy prior that the two models were equally likely, these probabilities were used to weight the contribution of V_s calculated under each model to a model-averaged estimate of V_s as below:

$$V_s = \left(0.5 * \text{HC pvalue} * {V_s}^{HC} + 0.5 * (1 - \text{HC pvalue}) * {V_s}^{G}\right) + (0.5 * \text{Gaussian pvalue} * {V_s}^{G} + 0.5 * (1 - \text{Gaussian pvalue}) * {V_s}^{HC})$$

The same procedure was repeated with the stochastic House of Cards (SHC) model to generate a model averaged estimate between the SHC and Gaussian models.

Genomic correlates and predictors of V_s

To test whether the degree of selective constraint predicted essentiality for each gene, we parameterized a normal distribution with the first central moment defined by the median of

the uncertainty distribution for mean population expression level and the second central moment defined by the median of the posterior distribution for HC V_s. The cumulative density function of this distribution was evaluated at 0 as a hard selective bound on expression level, and used as a logistic predictor for independently gathered data on essentiality in rich media(Seringhaus et al. 2006).

Diverse other genomic correlates tested for correlation with V_s included GC content, codon adaptation index, number of effective codons, rate of amino acid replacement, rate of amino acid substitution, protein length, recombination rate, hydropathicity, number of transmembrane helices, SGA scores, and metabolic network betweenness. Logistic regression was performed to identify relationships between binary variables presence of TATA box and gene essentiality and posterior median V_s.

Comparison with other phenotypes

To compare V_s estimates for gene expression to measures of stabilizing selection previously reported for morphological and phenological traits, we drew on values of the standardized selection gradients (γ) calculated as partial regressions of fitnesses on trait values (Lande and Arnold 1983) . Assuming a nor-optimal fitness distribution and using the approximation $\frac{V_s}{V_p} = -\frac{1}{2\gamma}$ (Johnson and Barton 2005), the magnitudes of stabilizing selection inferred for yeast gene expression traits were contrasted with all significant estimates of selection on phenological and morphological traits in a recent meta-analysis (Kingsolver et al. 2001; Kingsolver and Diamond 2011).

This comparison is a rough approximation for two reasons. Our analysis of genome-wide gene expression might be expected to overestimate the strength of selection on gene expression for any single gene as it does not partition the effect of selection into direct and indirect responses to selection as analyses of the phenological and morphological traits do (Lande and Arnold 1983). It should also be noted that relatively few statistically significant estimates of stabilizing selection in natural populations have been reported (Kingsolver et al. 2001; Kingsolver and Diamond 2011; Kingsolver et al. 2012), at least in part because the scaling of γ leads to a steep boundary effect for significance dependent on differences in fitness that may be difficult to measure with sufficient power. Additionally, experimental design may be biased to capture fast evolving traits with an a priori expectation of directional evolution (Kingsolver and Diamond 2011). For these reasons, the true distribution of strengths of stabilizing selection on organismal phenotypes may not be fully captured here. Nonetheless, we find this comparison of stabilizing selection on different levels of phenotypes useful in summarizing the current evidence and suggesting patterns for informing further investigations.

Supplementary Tables

Table S1: Genome-wide distribution of V_s values inferred under the House of Cards and Gaussian models using *n* based on the number of eQTLs identified for each gene. The genome-wide distribution is composed of the median values of the gene-specific uncertainty distributions.

	S. cerevisiae			***D. melanogaster* – European**			***D. melanogaster* - African**			***C. elegans***		
	Median	*0.05*	*0.95*	*Median*	*0.05*	*0.95*	*Median*	*0.05*	*0.95*	*Median*	*0.05*	*0.95*
V_s^{HC}	1324	443	4136	1358	231	3433	1259	233	4789	1078	369	7889
V_s^{G}	0.4	0.1	1.9	75	1.1	802	65	0.9	1668	3.2	0.3	219

Table S2: Proportion of total genes by species for which the Gaussian model could be conclusively rejected based on the probability of $20\mu V_s < \alpha^2$ using different models to estimate the quantity V_s. This analysis uses n based on the number of eQTLs identified for each gene. Drawing from exponential distributions with mean values between 1 and 500 produce qualitatively similar results.

	S. cerevisiae (n=3405)	***C. elegans*** (n=930)	***D. melanogaster*** (n=563)
V_s^{HC}	15%	74%	11%
V_s^{G}	96%	99%	69%
Model-averaged V_s^{HC} and V_s^{G}	15%	76%	11%

Supplementary Figures

Figure S1: Scatterplot of HC by SHC V_s values for *S. cerevisiae* (orange diamonds), *D. melanogaster* (green squares), and *C. elegans* (blue triangles). Values plotted are the gene-specific medians of the posterior distributions. This analysis uses *n* based on the number of eQTLs identified for each gene. Linear regressions are represented as solid lines. For *S. cerevisiae*, y = 1.0 x + 0.84, R^2 = 0.99; for *D. melanogaster*, y = 1.0x -17.4; R^2 = 0.99; for *C. elegans*; y = 1.3x -361, R^2 = 0.98

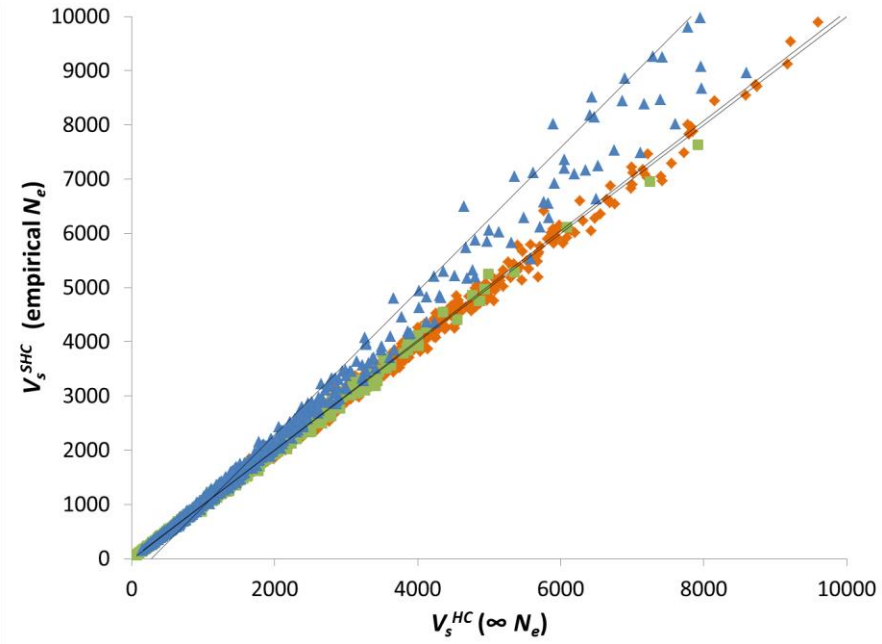
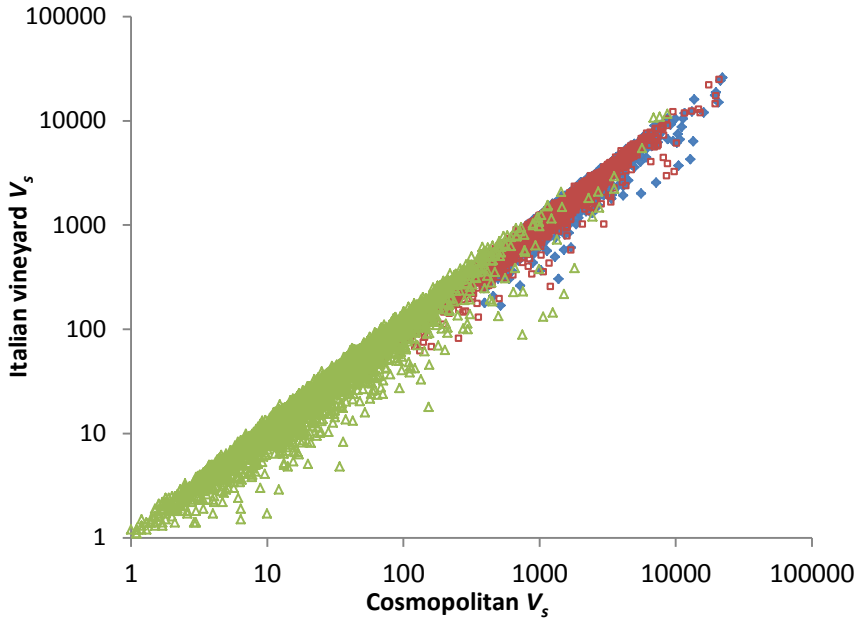


Figure S2: Influence of population choice on inference.

(a) Scatterplot of V_s values inferred when confining analyses to a subset of eight yeast isolates sampled from an Italian vineyard in consecutive years by V_s values inferred for the total population. The total yeast population includes two Pennsylvanian oak isolates and lab strain S288c. All estimates were scaled by S288C. Points plotted are gene-specific medians of uncertainty distributions for V_s^{HC} (blue triangles), V_s^{SHC} (red squares), and V_s^{G} (green triangles). This analysis uses *n* based on the number of eQTLs identified for each gene. Linear regressions of Italian on cosmopolitan V_s values without log transformation yielded y = 0.96x + 78, R^2 = 0.93 for V_s^{HC}; y = 0.99x + 20; R^2 = 0.95 for V_s^{SHC}; and y = 1.17x – 9.4, R^2 = 0.91 for V_s^{G}.

(b) Frequency distribution of V_s^{HC} values for two different populations of *D. melanogaster*. Population variation data for estimating V_g was available for *D. melanogaster* flies from natural populations in both the Netherlands (solid bars) and Zimbabwe (striped bars)(Hutter et al. 2008). Values plotted are the gene-specific medians of the V_s^{HC} posterior distributions using *n* based on the number of eQTLs identified for each gene. While individual genes differed in their V_s^{HC} values across populations from the Netherlands and Zimbabwe, the overall distribution of magnitudes for V_s^{HC} values estimated was highly similar.

a.


100000
10000
1000
100
10
1
Italian vineyard V_s
1
10
100
1000
10000
100000
Cosmopolitan V_s

b.

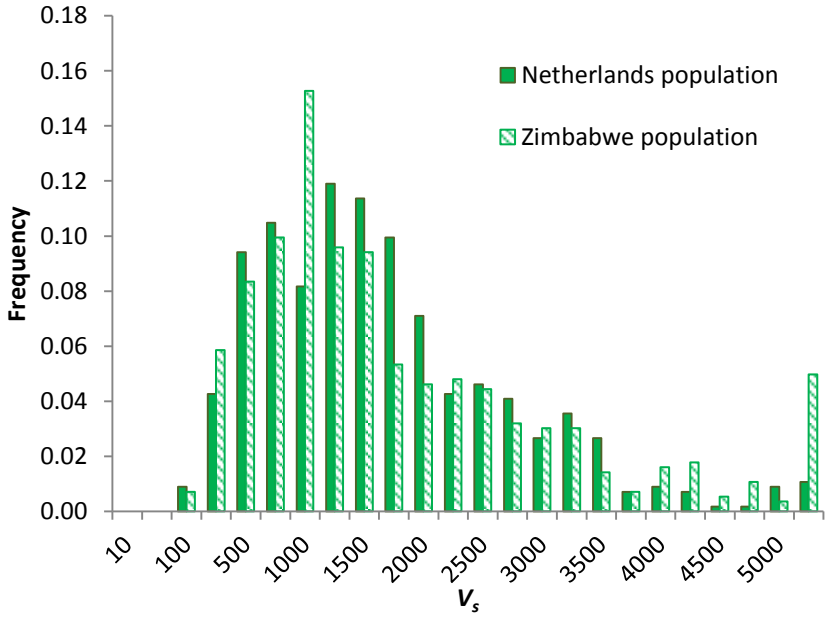
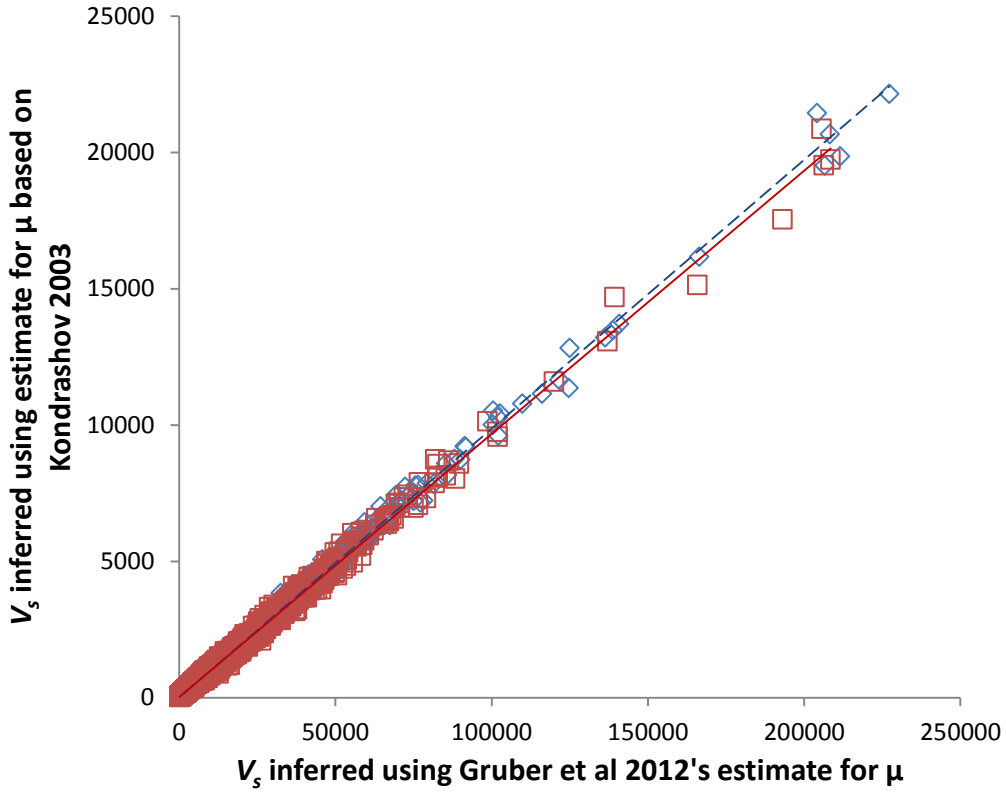

0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Frequency
Netherlands population
Zimbabwe population
10
100
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
V_s

Figure S3: Scatterplot of yeast V_s values inferred using two different sources for estimating the phenotypic per locus mutation rate μ influencing gene expression. The X axis represents the median gene-specific V_s values inferred when using a direct empirical point estimate of μ based on yeast gene TDH3 (Gruber et al. 2012). The Y axis represents the median gene-specific V_s values inferred using empirical data gathered for human Mendelian diseases phenotypes (Kondrashov 2003) to define a gamma distribution characterizing μ. This analysis uses *n* based on the number of eQTLs identified for each gene. Both the V_s^{HC} (blue diamonds) and V_s^{SHC} (red squares) values and linear regressions are plotted. For V_s^{HC}, y = 0.098x + 11.8, R^2 = 0.997; for V_s^{SHC}, y = 0.097x + 17.8, R^2 = 0.996.



Supplemental References

Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. 2005. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. Nature Genetics 37: 544-548. doi: 10.1038/ng1554

Fay J, McCullough H, Sniegowski P, Eisen M. 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. Genome Biology 5.

Fay JC, Benavides JA. 2005. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. PLoS Genetics 1: 66-71. doi: 10.1371/journal.pgen.0010005

Gruber JD, Vogel K, Kalay G, Wittkopp PJ. 2012. Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. Plos Genetics 8. doi: e100249710.1371/journal.pgen.1002497

Hoeting JA, Madigan D, Raftery AE, Volinsky CT. 1999. Bayesian model averaging: A tutorial. Statistical Science 14: 382-401.

Hutter S, Saminadin-Peter SS, Stephan W, Parsch J. 2008. Gene expression variation in African and European populations of Drosophila melanogaster. Genome Biology 9: R12.11-R12.15.

Johnson T, Barton N. 2005. Theoretical models of selection and mutation on quantitative traits. Philosophical Transactions of the Royal Society B-Biological Sciences 360: 1411-1425.

Kingsolver JG, Diamond SE. 2011. Phenotypic Selection in Natural Populations: What Limits Directional Selection? American Naturalist 177. doi: 10.1086/658341

Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM. 2012. Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. Evolutionary Ecology 26: 1101-1118. doi: 10.1007/s10682-012-9563-5

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P. 2001. The strength of phenotypic selection in natural populations. American Naturalist 157. doi: 10.1086/319193

Kondrashov AS. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Human Mutation 21: 12-27. doi: 10.1002/humu.10147

Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. Evolution 37: 1210-1226. doi: 10.2307/2408842

Seringhaus M, Paccanaro A, Borneman A, Snyder M, Gerstein M. 2006. Predicting essential genes in fungal genomes. Genome Research 16. doi: 10.1101/gr.5144106