Empirical measures of mutational effects define neutral models of regulatory evolution in *Saccharomyces cerevisiae*

Andrea Hodgins-Davis[a], Fabien Duveau[a], Elizabeth A. Walker[a], and Patricia J. Wittkopp[a,b,1]

[a]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109; and [b]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109

Edited by Michael Lynch, Arizona State University, Tempe, AZ, and approved September 4, 2019 (received for review February 16, 2019)

Understanding how phenotypes evolve requires disentangling the effects of mutation generating new variation from the effects of selection filtering it. Tests for selection frequently assume that mutation introduces phenotypic variation symmetrically around the population mean, yet few studies have tested this assumption by deeply sampling the distributions of mutational effects for particular traits. Here, we examine distributions of mutational effects for gene expression in the budding yeast *Saccharomyces cerevisiae* by measuring the effects of thousands of point mutations introduced randomly throughout the genome. We find that the distributions of mutational effects differ for the 10 genes surveyed and are inconsistent with normality. For example, all 10 distributions of mutational effects included more mutations with large effects than expected for normally distributed phenotypes. In addition, some genes also showed asymmetries in their distribution of mutational effects, with new mutations more likely to increase than decrease the gene's expression or vice versa. Neutral models of regulatory evolution that take these empirically determined distributions into account suggest that neutral processes may explain more expression variation within natural populations than currently appreciated.

gene expression | transcription | mutation | quantitative genetics

Variation in gene expression is widespread within and between species (1, 2). This variation reflects the joint action of mutation introducing new phenotypic variation and selection filtering variants based on their fitness effects. When genes exhibit differences in the rate at which they accumulate expression variation over time, selection pressure that varies among genes is often invoked to explain variability in expression divergence (3, 4). However, variability in the effects of new mutations on gene expression may also affect evolutionary outcomes by biasing the variety of expression phenotypes available for selection and drift to act upon (5, 6). Identifying such biases in mutational effects is challenging because the effects of mutation are confounded with the effects of selection in natural populations.

One way to isolate the effects of new mutations on gene expression is to perform a mutation accumulation (MA) experiment, which allows new mutations to accumulate in the near absence of natural selection (7). By comparing expression among evolved lines at the end of a period of MA, such experiments have been used to estimate mutational variance (V_M) for gene expression, which is the variance in a gene's expression due to spontaneous mutations each generation. V_M for gene expression differs among genes (8) and has been estimated genome-wide for model organisms including budding yeast (9), nematode worms (10, 11), and Drosophilid flies (12–15). Variation in expression V_M observed among genes in these studies reveals differences in the potential for a gene's expression to change over time. For example, in yeast, genes with more transcriptional regulators (as estimated from transcriptional profiling of gene deletion strains) tended to have higher V_M for expression than genes with fewer transcriptional regulators (9), suggesting that differences among regulatory networks can influence changes in gene expression due to new mutations.

Although these MA studies offer a global view of transcriptional changes across the genome, they provide a very limited view of the distribution of mutational effects for any single gene because the number of spontaneous mutations sampled in each study is low. For example, the 12.1-Mb yeast genome has a point mutation rate in the range of 10^{-9} to 10^{-10} per base pair per cell division, suggesting that only 4 point mutations occur on average every 1,000 cell divisions. Consequently, even ambitious MA experiments that capture 2,000 to 5,000 generations of spontaneous mutations are expected to survey fewer than 2 dozen point mutations per line. Mutagenesis studies, by contrast, can sample mutations affecting expression of a given gene much more deeply, typically trading off breadth of information across the genome for more focused and comprehensive descriptions of distributions of mutational effects for single genes. Thus far, such single-gene studies have focused primarily on mutations in *cis*-acting sequences controlling a gene's expression, such as promoters (16). For example, massively parallel reporter gene approaches have been used to describe the effects of thousands of mutations in promoters and enhancers on gene expression from diverse organisms including viruses, bacteria, yeast, and metazoans (e.g.,

Significance

New mutations tend to arise randomly throughout the genome, but their phenotypic effects are often not random. This disconnect results from interactions among genes that define the genotype–phenotype map. The structure of this map is poorly known and different for each trait, making it challenging to predict the distribution of mutational effects for specific phenotypes. Empirical measures of the distribution of mutational effects are thus necessary to understand how traits can change in the absence of natural selection. In this work, we define such distributions for expression of 10 genes in *Saccharomyces cerevisiae* and show that they predict greater neutral expression divergence than commonly used models of phenotypic evolution.

Author contributions: A.H.D., F.D., and P.J.W. designed research; A.H.D. and E.A.W. performed research; A.H.D. and F.D. contributed new reagents/analytic tools; A.H.D. analyzed data; and A.H.D., F.D., and P.J.W. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: All raw flow cytometry data reported in this paper have been deposited on FlowRepository, http://flowrepository.org/ (records FR-FCM-ZYUW and FR-FCM-ZZNR). Metadata and processed results have been deposited on the Deep Blue Data repository, https://deepblue.lib.umich.edu (DOI: 10.7302/0dvr-k169).

[1]To whom correspondence may be addressed. Email: wittkopp@umich.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1902823116/-/DCSupplemental.

First published September 30, 2019.

Downloaded at Palestinian Territory, occupied on December 23, 2021

www.manaraa.com

refs. 17–20). However, these *cis*-acting sequences are only one part of the mutational target for gene expression (21). Regions of the genome encoding or regulating *trans*-acting factors that interact with *cis*-acting sequences, either directly or indirectly, can also harbor mutations that affect gene expression (22). This *trans*-mutational target size is expected to be much larger than the *cis*-mutational target size (23) and can show different biases in the effects on expression (24).

Here, we use genome-wide mutagenesis to deeply sample and compare the effects of new mutations on expression of 10 focal genes. We observe differences in the distributions of mutational effects among these genes that are only partially captured by quantifying variance of mutational distributions (V_M). In particular, we observed differences in higher moments of these distributions, including the extent of asymmetry described by skewness and the frequency of mutations with extreme effects on expression related to kurtosis. Consistent with these observations, we find that all 10 distributions of mutational effects for gene expression are nonnormal with heavy tails (i.e., they contain more extreme events than a normal distribution). By using these empirically determined distributions of mutational effects to parametrize neutral models of gene expression evolution, we show that dramatic differences in expression divergence can occur among genes even in the absence of selection. In other words, we find that distributions of mutational effects for gene expression are 1) more complex than frequently assumed, 2) different among genes, and

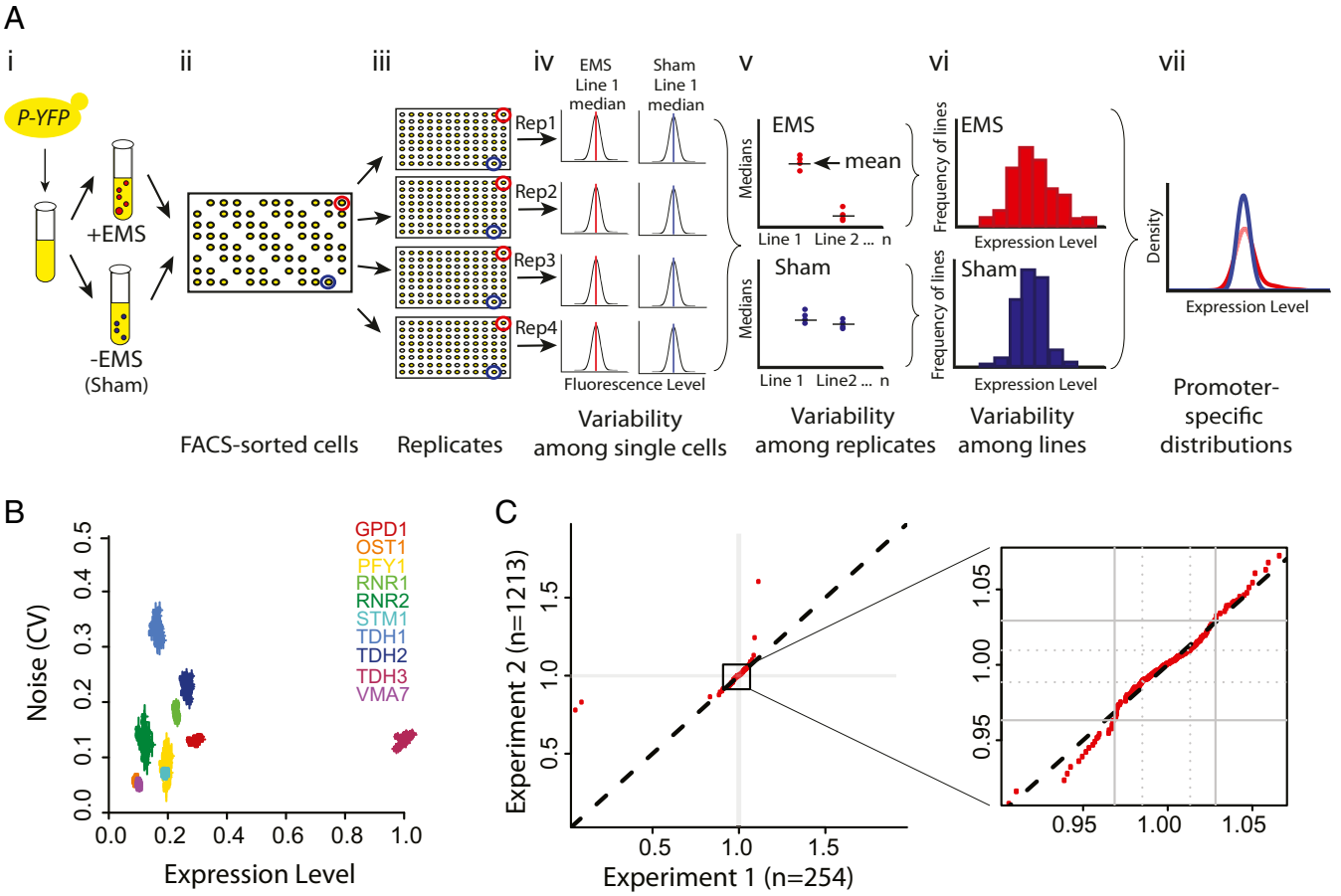

Fig. 1. Experimental overview. (*A*) A schematic describing the experimental design is shown: (*i*) cells containing a promoter sequence (*P*) driving expression of the yellow fluorescent protein (*YFP*) coding sequence were grown to saturation and split into samples for mutagenesis and sham treatments. (*ii*) Following these treatments, single cells were deposited on YPD agar plates using fluorescence-activated cell sorting (FACS) in a 96-well plate format, leaving empty spaces for controls. At least 8 plates were collected per promoter, with each including EMS-treated cells (n = 30/plate) and sham-treated cells (n = 8/plate for both promoter-matched and P_{TDH3}-*YFP* shams) as well as wild-type control samples (n = 24/plate) and nonfluorescent control lines (n = 1/plate) added to the empty spaces. (*iii*) Each of these plates was used to inoculate liquid cultures in 96-well plates in quadruplicate, with fluorescence in the resulting cells scored using flow cytometry. (*iv*) YFP expression was estimated (BD Accuri C6) for each cell in each sample as the ratio of fluorescence (log FL1-A: 488-nm laser, 530/30-nm filter) and cell size (log FSC-A). YFP fluorescence was converted to estimated mRNA level, adjusting for known nonlinearity between YFP fluorescence and mRNA abundance (78). Median YFP fluorescence was then taken as the expression level for each sample. (*v*) For each EMS-treated and sham-treated line, expression level was calculated as the mean of the median expression levels from the 4 replicate populations. (*vi*) Variability in expression among lines isolated from the EMS-treated and sham-treated populations were used to estimate promoter-specific distributions (*vii*) reflecting mutational and environmental variance, respectively. (*B*) Median expression levels (x axis) and expression noise (y axis) are shown for 40 independent unmutagenized (sham-treated) lines per promoter. Each data point represents the mean of 4 replicate flow cytometry measurements per line scaled relative to the activity of the *TDH3* promoter. Expression noise was measured as the coefficient of variation (CV) among cells measured within each replicate, and the mean CV was calculated from the 4 replicates. Error bars represent 95% confidence intervals for each line. (*C*) Quantile–quantile plot comparing the distribution of mutational effects for 2 independent mutagenesis experiments for the *TDH3* promoter (x axis: experiment 1, n = 254; y axis: experiment 2, n = 1,213) are shown. The second panel enlarges the area corresponding to the central 95% density of both experiments (solid lines: 10th and 90th quantiles; dotted lines: 25th and 75th quantiles). The dashed line on the diagonal represents the hypothesis that the 2 samples are drawn from the same distribution. A nonparametric AD test fails to reject the null hypothesis that these 2 samples come from a common population (AD criterion = 0.76056; P = 0.29).

Downloaded at Palestinian Territory, occupied on December 23, 2021
www.manaraa.com

3) able to introduce biases in the direction of neutral evolution. These observations suggest that failing to account for mutational biases may underestimate the role of neutral evolution in expression divergence.

Results and Discussion

To compare the effects of mutations on gene expression driven by promoters from different genes, we selected 10 promoters from *Saccharomyces cerevisiae*: *GPD1*, *OST1*, *PFY1*, *RNR1*, *RNR2*, *STM1*, *TDH1*, *TDH2*, *TDH3*, and *VMA7*. These promoters vary for properties previously hypothesized to affect the mutability of gene expression, including expression noise (25), nucleosome occupancy (25), and number of mismatches to a canonical TATA box (9) (*SI Appendix*, Table S1). This set of genes includes 3 paralogs, *TDH1*, *TDH2*, and *TDH3*, and 2 genes acting in the same molecular complex, *RNR1* and *RNR2*. All of these promoters drive expression at levels that can be reliably detected by flow cytometry. For each gene, we cloned the promoter sequence upstream of the coding sequence of a yellow fluorescent protein (YFP) and inserted the resulting reporter gene into the *S. cerevisiae* genome at the *ho* locus. YFP expression level from these reporter genes is therefore expected to measure effects of mutations in the *cis*-acting promoter sequence as well as all *trans*-acting regulators of the gene from which the promoter was derived.

To assess the impact of new mutations on expression driven by each promoter, we exposed cells carrying each reporter gene to a low dose of the chemical mutagen ethyl methanesulfonate (EMS) (Fig. 1*A*). We also subjected cells with each promoter to the same mutagenesis protocol except for the exposure to EMS to create a control ("sham-treated") population that captured nongenetic sources of variability. EMS introduces G→A and C→T point mutations (26). While these mutations are a subset of the types of changes that arise spontaneously, they are the most common type of point mutation observed in MA lines (27, 28) and the most common type of single-nucleotide polymorphism segregating in natural populations of *S. cerevisiae* (29, 30). Using a canvanine resistance assay with a known mutation rate (31, 32), we determined that the EMS conditions used introduced ~29 mutations per cell (95% percentiles: 24 to 39). This same level of mutagenesis was used previously to characterize effects of new mutations on YFP expression driven by the *TDH3* promoter (P_{TDH3}-*YFP*) (24, 32). Genetic mapping has shown that mutants isolated under these conditions typically carried no more than 1 mutation causing a measurable change in P_{TDH3}-*YFP* expression (33); thus, most mutations introduced are expected to have no effect on expression of the focal gene.

From both the EMS- and sham-treated populations, we used fluorescence-activated cell sorting (FACS) to isolate single cells randomly with respect to the YFP fluorescence level (Fig. 1*A*). For each promoter, at least 240 mutagenized cells and at least 64 sham-treated cells were sorted individually onto solid YPD agar plates in a 96-well format. After allowing these sorted single cells to grow into colonies, establishing a distinct "line" from each sorted cell, 4 replicate populations were inoculated for each line (Fig. 1*A*). YFP fluorescence levels were then estimated for at least ~12,000 events captured with flow cytometry from each replicate population (*SI Appendix*, Table S2 and refs. 34 and 35). Ultimately, we established 148 to 254 lines for each promoter from the mutagenized population (median, 214) as well as 44 to 62 lines for each promoter from each sham-treated population (median, 55). After correcting for variation in cell size, the median YFP expression level within each replicate was taken as a point estimate of that sample's average expression level (Fig. 1*A* and ref. 36). The expression level for each line was then defined as the mean of the median values from the 4 replicates (Fig. 1*A*). This same procedure was used to estimate YFP expression levels from similar data in prior work (37).

The 44 to 62 lines established from each sham-treated population were expected to be genetically identical (barring spontaneous mutations); thus, variation in the expression level among the sham-treated lines for each promoter was used to estimate the nongenetic deviation among individuals separated into lines derived from single cells, akin to quantitative genetic environmental variance (V_E). Comparing expression of lines derived from the sham-treated population for each promoter shows differences among promoters in average expression level, expression noise, and environmental variance (Fig. 1*B*). Each line established from the mutagenized population is expected to carry a distinct set of new mutations; thus, variation in the expression level among the 148 to 254 mutagenized lines for each promoter reflects a combination of genetic differences among lines and environmental variance, with the environmental variance expected to be equal to that observed among the lines isolated from the sham-treated population. Comparing the distribution of mutational effects observed for the 254 mutagenized lines affecting expression of the *TDH3* reporter gene to the distribution of mutational effects inferred previously using an independent collection of >1,200

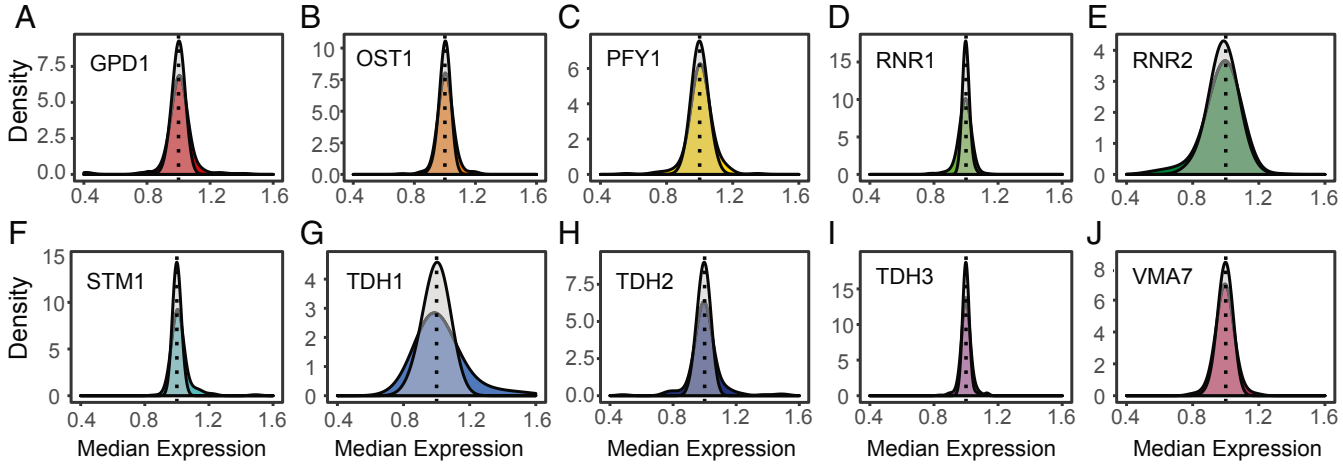

Fig. 2. Distributions of expression levels observed for mutagenized and unmutagenized lines. (*A–J*) Distributions of expression among mutagenized (colors) and sham-treated lines (overlaid gray transparency) for each promoter. Expression levels (*x* axis) in both mutant and sham populations are expressed relative to the median of each wild-type expression level (dotted line). The frequency of phenotypes in each population is summarized by promoter as a smoothed density curve (*y* axis). Data for *TDH3* in *I* were previously published in ref. 24. Sample sizes are listed in *SI Appendix*, Table S3.

Downloaded at Palestinian Territory, occupied on December 23, 2021
www.manaraa.com

mutagenized lines carrying the same *TDH3* reporter gene (24) (Fig. 1*C*) suggested that the sample sizes used in this study are sufficient to provide reasonable approximations of the underlying mutational distributions.

Distributions of Mutational Effects Differ in Skewness, Kurtosis, and Dispersion. For all promoters, we observed that the distribution of expression levels for sham-treated lines was symmetrical (Fig. 2, gray distributions), indicating that nongenetic (environmental) sources of variation were equally likely to increase and decrease expression relative to the median. The distribution of expression levels for EMS-treated lines was asymmetric, however, for some promoters (Fig. 2, colored distributions; *SI Appendix*, Table S3), suggesting that mutations have biased effects on the expression driven by these promoters. Specifically, 3 promoters exhibited statistically significant departures from symmetry resulting from differences in the magnitude of increased or decreased expression in mutagenized lines relative to the median of the sham-treated lines: Mutagenized lines carrying the *TDH1* and *STM1* promoters showed larger increases than decreases in expression (*SI Appendix*, Table S3: permutation test, *TDH1*: $P = 0.006$; *STM1*: $P = 0.031$), and mutagenized lines carrying the *RNR2* promoter showed larger decreases than increases in expression ($P = 0.013$).

Before comparing the distributions of mutational effects derived from the mutagenized lines among promoters, we took into account differences in the environmental variance among promoters revealed by differences in the distributions of expression levels for sham-treated lines among promoters. To do so, we converted the expression level measured for each mutagenized line into a *Z* score (*SI Appendix*, Fig. S1) by subtracting the median expression level of sham-treated lines with the same promoter and then dividing by the SD of expression levels for these sham-treated lines. In this *Z*-score scale, 1 unit corresponds to a change in expression equivalent to 1 SD in the population of sham-treated lines for that promoter. Variance of these *Z* scores is equivalent to mutational heritability calculated by dividing the (mutational) variance observed among the mutagenized lines by the (environmental) variance estimated by the sham-treated lines; mutational heritability is the metric used to compare the effects of mutations on different traits (38). For all promoters, we found that at least 60% of mutagenized lines (range, 60 to 86%) had expression *Z* scores within 2 SDs of the sham median, indicating that most mutagenized lines showed expression levels similar to those occurring due to environmental variance (Fig. 3*A* and *SI Appendix*, Table S4).

The distributions of *Z* scores for EMS-treated lines violated normality for all 10 promoters (Shapiro–Wilks test; *SI Appendix*, Table S5). To more fully describe the shapes of these distributions, we calculated robust summary statistics that are less sensitive to outliers than the traditional measures used to estimate moments of a distribution (39–41) (*SI Appendix*, Figs. S2–S5). Comparing the shapes of distributions of mutational effects among promoters (Fig. 3*A* and *SI Appendix*, Fig. S6), we observed differences in the centrality (median), dispersion (median-averaged deviation or MAD), skewness (medcouple or MC), and relative frequency of lines with extreme effects (left/right medcouple or LMC/RMC; *SI Appendix*, Table S4). A principal-component analysis of summary statistics describing these properties found that the first principal component explained 36% of the variance among promoters and primarily captured skewness and the frequency of extreme increases in expression (*SI Appendix*, Fig. S6 *A* and *D*). A second, independent, principal component explained 30% of the variance and was strongly influenced by median and dispersion (*SI Appendix*, Fig. S6 *B* and *E*). Finally, the third principal component explained 24% of the variance and was influenced by both extreme decreases in expression and dispersion (*SI Appendix*, Fig. S6 *C* and *F*). Because differences in symmetry among promoters dominated these contrasts, we chose to more directly examine skewness for a range of effect sizes using quantile–quantile plots comparing the magnitude of increases to the magnitude of decreases moving away from the median for each promoter (Fig. 3*B*). By illustrating biases in the direction and magnitude of mutational effects as departures from the 45° line, these plots highlight differences among promoters such as the asymmetries described above for *STM1*, *RNR1*, *TDH1*, and *RNR2*.

Directly comparing the distributions of *Z* scores between promoters, we found 13 of 45 pairwise comparisons had distributions of mutational effects that differed significantly between promoters (Anderson–Darling test [AD], $P < 0.05$ with Benjamini–Hochberg [BH] correction for multiple tests; *SI Appendix*, Fig. S7 and Table S6). Seven of these 13 cases involved *RNR1*, which possessed an especially unique distribution of mutational effects due to a wide dispersion of *Z* scores among mutagenized strains and an overall negative skew, including a small decrease in median compared to its sham population. The distribution of mutational effects for *STM1* was significantly different from 5 other promoters including *RNR1*, exhibiting biases in the opposite direction from the *RNR1* distribution: *STM1* also showed broad dispersion, but exhibited an overall positive skew with more large effect increases in expression than any other promoter, including a small increase in median compared to its sham population. Other distributions of mutational effects that showed pairwise differences from more than one other promoter included *TDH1* and *RNR2*, which were significantly different from each other and 2 other promoter distributions (*TDH1*: *RNR1*, *VMA7*; *RNR2*: *RNR1*, *STM1*). Like *STM1*, *TDH1* exhibited a positive overall skew with more density in the right tail of its distribution of mutational effects, but without the right shift in median. *RNR2* was distinct in showing the most humped (also known as platykurtic) distribution

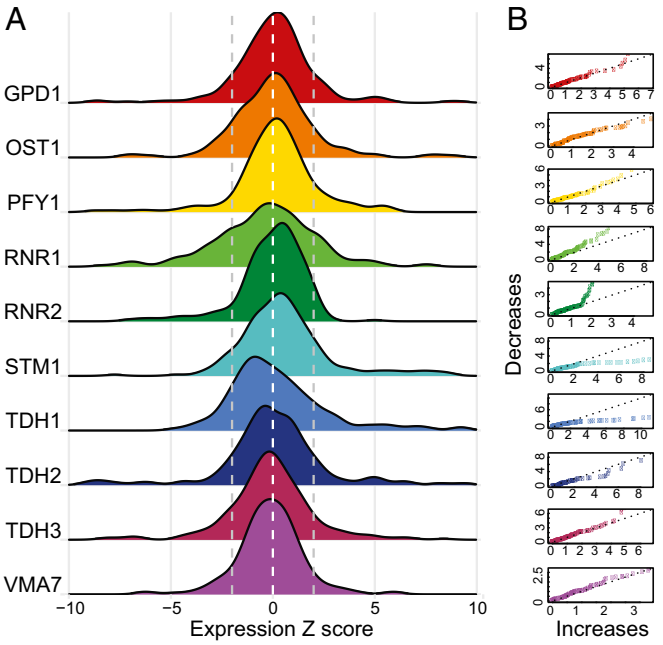

Fig. 3. Distributions of mutational effects. (*A*) Stacked distributions of mutational effects among mutagenized populations. Mutational effects (*x* axis) are expressed as *Z* scores. Dashed lines note the sham median and 2 SDs above and below the sham median. (*B*) The relative frequency of expression changes for a continuous range of magnitudes above and below the sham median are represented as quantile–quantile plots of the magnitude of increases in expression (*x* axis) by the magnitude of decreases in expression (*y* axis) for each promoter compared to the median sham phenotype. The dashed line on the diagonal represents the hypothesis that mutational effects generate increases and decreases from the sham median with equal frequency at all effect sizes.

Downloaded at Palestinian Territory, occupied on December 23, 2021
www.manaraa.com

with the least density in the extreme tails. This *RNR2* distribution also showed a slight shift in median toward increased expression despite an overall negative skew.

Relationships between Promoter Properties and Distributions of Mutational Effects. As described above, the promoters included in this study were chosen because they vary for properties hypothesized to influence the mutability of gene expression. To determine whether any of these properties might explain the differences in distributions of mutational effects that we observed, we tested for evidence of a significant relationship between the robust summary statistics describing the empirical distributions of mutational effects and the following 7 gene properties: 1) expression level of the native gene, 2) expression noise for the native gene, 3) presence of a canonical TATA box in the gene's promoter, 4) number of *trans*-regulatory factors annotated in YEASTRACT for the gene, 5) density of nucleosome occupancy in the gene's promoter region, 6) presence of a duplicate gene in the yeast genome, and 7) fitness of strains homozygous for a deletion of the gene in rich media. We observed no statistically significant relationship between either dispersion (breadth) measured as MAD or skewness measured as MC and any of the 7 properties tested after correction for multiple tests (*SI Appendix*, Fig. S8). We also tested whether the nongenetic variability of each promoter observed among the sham-treated lines could predict mutational variance, as suggested by prior work (9), but did not observe a significant correlation between the variance in Z scores among EMS-treated lines and the variance of expression levels in the corresponding sham-treated lines (Spearman's rho = −0.27, $P = 0.448$). For all of these analyses, our power to detect significant effects was limited by the number of genes analyzed. Future studies deeply sampling the effects of mutations on many more genes are needed to better understand how properties of promoters, or the regulatory networks they are embedded in, affect gene-specific distributions of mutational effects for gene expression.

Predicting Neutral Expression Divergence Using Distributions of Mutational Effects. In the absence of empirical data describing the distribution of mutational effects for a specific trait, tests for selection often make the simplifying assumption that distributions of mutational effects are symmetric or normally distributed (e.g., ref. 42). This assumption is based on the idea that quantitative traits are generally controlled by many loci with small effects (43). If traits are controlled by relatively few loci and/or loci of large effect, as sometimes seems to be the case for gene expression (22, 44), the distribution of mutational effects may be likely to violate normality. Our study supports this observation for gene expression phenotypes, and studies of mutational effects for morphological traits (largely in *Drosophila*) have also tended to produce nonnormal (leptokurtic) distributions of mutational effects with heavy tails (45–53), suggesting they might have similar genetic architecture. Theoretical work has shown that ignoring nonnormality of distributions of mutational effects can cause evolutionary models to produce misleading inferences (54–57), but the sparseness of empirical data describing distributions of mutational effects has limited our ability to assess the magnitude of errors caused by these assumptions. To address this knowledge gap, we used our empirical distributions of mutational effects to parametrize models of neutral regulatory evolution for the 10 genes examined, contrasting expression divergence predicted by simulations incorporating the full empirical distributions of mutational effects and Brownian motion models assuming mutational effects are normally distributed with a variance equal to the empirically observed mutational variance.

For each promoter, we simulated a population of 1,000 individuals with expression levels drawn randomly with replacement from the distribution of expression levels for that promoter's sham population. Each individual then had a probability of mutating equal to the spontaneous mutation rate observed in a *S. cerevisiae* MA study [1.67×10^{-10} per nucleotide per generation (27)], resulting in a new mutation arising in 2 individuals on average each generation. The effect of each mutation was determined by randomly sampling with replacement from the distribution of mutational effects for that promoter and multiplying this effect by the individual's original expression level, making the simplifying assumption that the distributions of mutational effects stay constant over time. This approach also assumes that the variation in expression among EMS-treated lines is attributable to a single causal mutation within each line, with no influence from other mutations or epistatic effects. We think that this assumption is reasonable because only one mutation was found to be responsible for changes in P_{TDH3}-*YFP* expression mapped in mutants generated with the same mutagenesis treatment (33), although the frequency of mutations affecting expression of other genes is expected to differ and remains unknown. One thousand simulated individuals were then randomly sampled with replacement to populate the next generation. This procedure was repeated for 50,000 generations, calculating mean expression level within the population at each generation. Five hundred independent simulations were run for each promoter to determine the variation in simulated mean expression levels at each generation.

At the end of 50,000 generations, expression divergence differed dramatically among promoters (Fig. 4). Promoters with a strong positive mutational skew in the distribution of mutational effects like *STM1* and *TDH1* exhibited large increases in median population expression levels across 500 independent evolutionary trajectories, while promoters with a strong negative mutational skew like *RNR1* and *RNR2* showed large declines in median population expression levels. Promoters with more symmetric mutational distributions, for example, *VMA7* and *OST1* (*SI Appendix*, Table S3), exhibited less expression divergence from the original expression level. The *TDH3* mutational distribution was also symmetric but included a few mutants with low expression relative to the rest of the population that caused large step-like decreases in expression when sampled; excluding the 5 lowest values or sampling from the larger collection of mutant phenotypes from ref. 24 resulted in much more symmetric evolution of *TDH3* expression (*SI Appendix*, Fig. S9). Among all 10 evolutionary trajectories, promoters with mutational distributions that had a greater skew or heavier weight in one tail diverged more, as expected: grand median expression at generation 50,000 was jointly predicted by skewness and weight in the extreme negative tail of the promoter-specific mutational distributions (grand median expression [log transformed to improve normality] at generation 50 K ~ MC + LMC, $F_{(2,7)} = 20.15$, $P = 0.001$). Similar results were observed using simulations with a population size of 100 instead of 1,000 individuals (*SI Appendix*, Fig. S9 *C–L*) and when using expression levels converted to Z scores (*SI Appendix*, Fig. S10).

We next simulated changes in expression expected for each gene under the more commonly used model of random walks in phenotype space described by Brownian motion. In this model, mutational effects were drawn from a normal distribution centered on the expression level of the unmutagenized promoter with variance equal to the variance observed among the mutagenized lines analyzed for that promoter. We again examined the population means of 500 independent simulations after 50,000 generations. We found that the Brownian motion simulations parametrized only with the empirical variance showed less overall divergence from the starting point than simulations drawing mutational effects from the full empirical distributions, although the extent of difference between the 2 models varied among promoters (Fig. 5, mean expression [log transformed] at generation 50 K ~ Promoter × Model Type: $F_{(19,9980)} = 616$, $P < 2.2 \times 10^{-16}$ with significant interaction identified by ANOVA F test, $P < 2.2 \times 10^{-16}$; *SI Appendix*, Table S7).

Downloaded at Palestinian Territory, occupied on December 23, 2021
www.manaraa.com

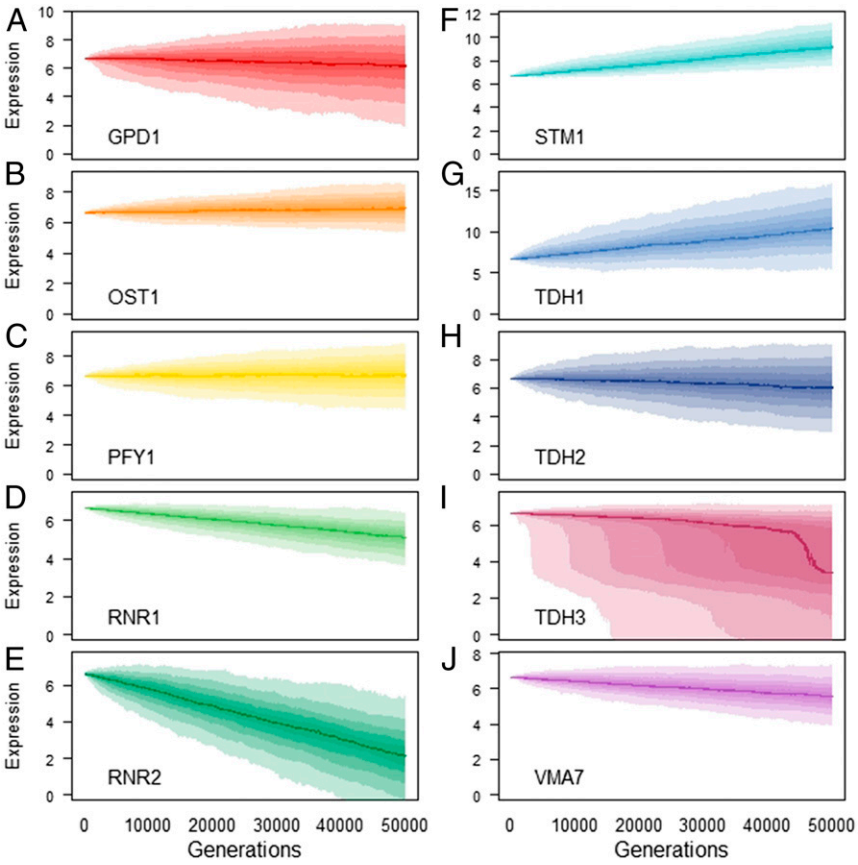

Fig. 4. Neutral changes in gene expression predicted by distributions of mutational effects. Evolutionary simulations based on empirical mutational distributions for populations of 1,000 individuals evolving in the absence of selection for 50,000 generations are shown for all 10 promoters (*A*–*J*). Population divergence at each generation (*x* axis) was summarized by the mean population phenotype represented on a $\log_2$ scale (*y* axis). Shaded areas represent the 95% credible intervals for the mean population phenotype at each generation among 500 independent evolutionary trajectories. The darkest line represents the grand median of all independent simulations, and lighter shading moving away from the grand median represents quantiles of replicate populations in increments of 10. All populations evolved from a starting population with a mean of $\log_2(100)$ corresponding to 6.64.

To determine how well empirically derived distributions of mutational effects might estimate neutral trait evolution, we compared properties of gene-specific distributions of mutational effects to levels of polymorphism seen for that gene among 22 natural isolates of *S. cerevisiae* (48) because polymorphism is often assumed to primarily reflect neutral processes (47). We observed a significant positive relationship between the degree of dispersion (MAD) in the mutational distribution and expression polymorphism measured as the expression variance among the 22 natural isolates [*SI Appendix*, Fig. S11, ExpVar ~ MAD, $F_{(1,8)} = 8.18$, $P = 0.021$]; however, this relationship was driven primarily by *TDH1*, which was an outlier for MAD with respect to the other promoters (*SI Appendix*, Fig. S4). Excluding *TDH1* reduced the strength of this correlation and resulted in a *P* value that was not statistically significant [$\text{ExpVar}_{\text{noTDH1}} \sim \text{MAD}_{\text{noTDH1}}$, $F_{(1,7)} = 1.122$, $P = 0.325$, dotted line]. Skewness of the distribution of mutational effects also failed to significantly predict polymorphism (ExpVar ~ MC, $F = 0.58$, $P = 0.47$); however, in both cases, we note that our power to detect such relationships is limited by the number of genes studied. Testing for relationships between the effects of mutation and polymorphism more robustly will require similarly deep sampling of mutational effects for many more genes in the yeast genome.

Modeling Distributions of Mutational Effects and the Evolution of Gene Expression. One of the benefits of assuming a normal distribution of mutational effects is that it simplifies modeling by allowing draws from a well-known continuous distribution rather than a collection of discrete empirical data points. We therefore sought to identify continuous probability distributions that reflect the shape of the observed empirical distributions of mutational effects better than normal distributions. Distributions of mutational effects for leaf traits in *Arabidopsis* have previously been described using the family of distributions known as LaPlace distributions (53), also known as double-exponential distributions, which have fatter tails than the normal distribution and can be specified in both symmetric (2-parameter) and asymmetric (3-parameter) forms. To determine whether LaPlace distributions fit distributions of mutational effects for gene expression better than normal distributions, we used maximum likelihood to optimize parameters for Gaussian, symmetric LaPlace, and asymmetric LaPlace distributions. Bayesian information criteria (BIC) were used to identify the best-fitting distribution for each promoter. For all 10 promoters, LaPlace distributions were better supported than a normal distribution for the observed distribution of mutational effects (*SI Appendix*, Table S8). The *VMA7* promoter exhibited similar levels of support for symmetric and asymmetric LaPlace distributions, whereas all other promoters were best described by asymmetric LaPlace distributions. These observations suggest that LaPlace distributions may provide more realistic distributions of mutational effects than normal distributions. They also encourage further investigation into models of regulatory evolution that relax the assumption that all loci have equal effects (54–56) and favor models of phenotypic evolution that allow for a high variance in mutational effects (58).

Conclusion

By studying the effects of thousands of new mutations on expression of individual genes, we have shown how distributions of mutational effects for gene expression differ among genes. Differences observed in the direction and magnitude of mutational effects suggest that some genes may exhibit underlying biases in the expression variation available to selection. In addition, large changes in gene expression were more common than predicted by a normal distribution. For most genes, a null model of neutral expression divergence based on sampling mutations from these

Downloaded at Palestinian Territory, occupied on December 23, 2021
www.manaraa.com

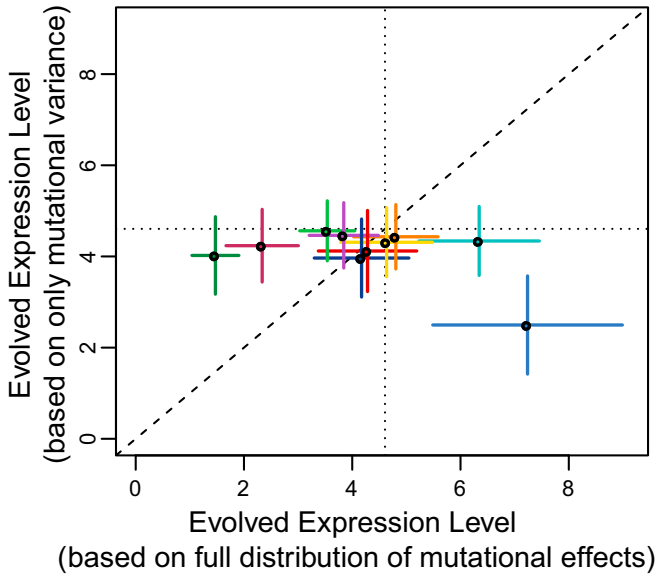

Fig. 5. Contrasting neutral models of regulatory evolution. Evolutionary divergence predicted after 50,000 generations by neutral simulations drawing mutational effects from the full empirical distribution of mutational effects (*x* axis) or a Brownian motion model drawing mutations from a normal distribution parametrized with the variance of the empirical mutational distributions (*y* axis). Points represent the grand median expression level among independent evolutionary trajectories for each gene at generation 50,000 under each model on a log scale. Error bars represent 95% credible intervals among 500 replicate populations.

distributions predicted greater expression divergence than Brownian motion models parametrized with mutational variance alone, suggesting that neutral evolution might explain more variability in gene expression within and between species than often assumed. Challenges for the future include 1) deeply characterizing the distribution of mutational effects for more genes, 2) measuring differences in the frequency of mutations affecting expression among genes, 3) determining how distributions of mutational effects vary among genetic backgrounds due to epistasis, and 4) identifying features of regulatory networks that can be used to predict a particular gene's propensity for mutations of a certain effect. Because gene expression is a critical step in the conversion of genotypes to phenotypes, addressing these issues will improve our understanding of the evolutionary processes that generate, maintain, and control variation in complex traits more generally.

Materials and Methods

More detailed information on the materials and methods used in this study are provided in *SI Appendix, Materials and Methods*.

Promoter Selection. Promoters from the *GPD1*, *OST1*, *PFY1*, *RNR1*, *RNR2*, *STM1*, *TDH1*, *TDH2*, *TDH3*, and *VMA7* genes were selected to represent a diverse range of properties expected to impact mutational variability (9, 18), including expression noise (59), presence of a TATA box motif (60), variation in nucleosome occupancy (61), mutational variance in MA studies, and fitness of homozygous gene deletion strains (62) (*SI Appendix*, Table S1). Maximizing sensitivity for flow cytometry required that all promoters drive relatively high expression, and promoters were thus selected from the 15% most highly expressed genes in *S. cerevisiae* (63).

Strain Creation. To assay promoter activity, a construct consisting of the promoter region of each focal gene (defined as the intergenic sequence between the start codon of the focal gene and next upstream gene) followed by the Venus YFP coding sequence, the CYC1 terminator, and an independently transcribed KanMX4 drug resistance was integrated at the HO locus of a BY4724-derivative strain, YPW1139, described in ref. 24. Constructs were generated through tailed PCR and transformed via homologous recombination into a strain with a ho::URA3-YFP allele. The genetic background of this strain carried the alleles RME1(ins-308A); TAO3(1493Q) (64) and SAL1; CAT5(91 M); MIP1(661T) (65), which decrease petite frequency relative to the alleles of the ancestral BY4724. Data reported include previously published results for *TDH3* (24) reanalyzed in a common framework with the 9 additional promoters reported here.

Mutagenesis. To sample the genome-wide effects of point mutations on promoter activity, we performed random mutagenesis of strains carrying all promoter constructs. Mutagenesis was executed as in ref. 24 using the chemical mutagen EMS, which introduces G/C to A/T point mutations randomly throughout the genome (see *SI Appendix* for details). Based on canvanine resistance assays performed for *P-TDH1-YFP*, we estimated that an average of ~29 mutations were introduced per cell with the EMS conditions used (95% percentiles: 24 to 39), consistent with refs. 24 and 32. Sham-treated controls including both a promoter-matched genotype and a *P-$_{TDH3}$-YFP* construct were maintained in parallel and treated identically with the exception of EMS exposure. Following mutagenesis, single cells from EMS- and sham-treated populations were isolated via FACS and recovered on YPD agar (2% dextrose, 1% yeast extract, 2% peptone, 2% agar) for 48 to 72 h at 30° C. A minimum of 8 plates were sorted per promoter construct with 30 mutagenized cells, 8 promoter-matched sham-treated cells, and 8 *P-TDH3-YFP* sham-treated cells interspersed among 24 control positions. Viability of isolated cells was significantly impacted by treatment condition, but not by the genetic background or mutagenesis batch [glm quasibinomial models: (Viability ~ Condition) vs. (Viability ~ Condition + Batch), *F* test, $F_{(47,55)} = 0.931$, $P = 0.5002$; (Viability ~ Condition) vs. (Viability ~ Condition × Batch), *F* test, $F_{(15,55)} = 0.5656$, $P = 0.9242$].

After the isolated cells were grown into colonies, the colonies were transferred from agar to liquid YPD and grown to saturation with shaking, typically 20 to 24 h at 30° C. These saturated cultures were then used 1) to preserve each line as a glycerol stock and 2) to initiate cultures for analyzing fluorescence. Cultures for analyzing fluorescence were spotted on YPD agar, and 20 to 24 plate control strains for estimating random experimental effects were interspersed into each row and column. After an additional 48- to 72-h growth, colonies were transferred to liquid YPD in 96-well deep-well culture plates, grown for 20 to 24 h to saturation with shaking, and scored for fluorescence. A minimum of 4 replicate assays were performed for each plate.

Phenotyping and Data Processing. To characterize promoter expression levels, YFP fluorescence driven by the promoter of interest was quantified for all sham-treated, EMS-treated, and plate control samples. Fluorescence data were collected on a BD Accuri C6 (488-nm laser and 530/30 optical filter) coupled with an IntelliCyt HyperCyt autosampler. Cultures were diluted in 1× PBS to ~10^6 cells per mL directly before scoring. A total of 48 samples was collected per instrument run with gentle vortexing to aerate and resuspend cells between runs. Samples analyzed sequentially within the same instrument run were distinguished by the Hyperview software (Intellicyt). Appropriate separation of samples was manually checked for every plate.

Using tools from the flowCore and flowClust libraries (66, 67) and custom scripts (*Access to Data and Analysis Scripts*), flow cytometry data were analyzed to remove noncellular debris, events where more than 1 cell passed the detector, extreme outliers in cell size or YFP, and correlation between cell size and YFP expression (*SI Appendix, Materials and Methods*). Additionally, because fixed photomultiplier tube (PMT) voltages on the Accuri C6 produce nonlinearity between fluorophore concentration and fluorescence intensity level (68), this study follows ref. 37 in using a standard curve determined by quantifying RNA abundance via pyrosequencing and fluorescence intensity via flow cytometry for the same samples to scale mRNA abundance estimates appropriately (*SI Appendix, Materials and Methods*). These procedures were performed on a single-cell basis for all events that passed quality control thresholds in each sample well. Individual samples were then summarized by calculating median YFP RNA abundance and coefficient of variation (CV) (estimated as MAD/median). Cell size was also summarized as median FSC and FSC MAD. A number of samples were excluded at this stage for phenotypes consistent with high levels of bacterial contamination (small cell size and no YFP expression) or contamination with *P-TDH3* sham controls (YFP expression at the median of the P_{TDH3}-*YFP* sham lines for a non P_{TDH3}-*YFP* sample). Any samples with fewer than 1,000 single cells passing quality control filters were excluded from analysis. The median and minimum number of cells analyzed per sample are listed by promoter in *SI Appendix*, Table S4.

To account for technical variability across plates, YFP mRNA abundance and FSC metrics for each sample were then normalized to remove random effects due to technical noise arising among instrument runs, plate row

Downloaded at Palestinian Territory, occupied on December 23, 2021
www.manaraa.com

position, or plate column position. The power to perform these normalizations came from inclusion of 20 to 24 control strains in each experimental 96-well plate. Initial experiments were performed using a *P-TDH3-YFP* construct in control positions as in previous work, but when contrasts showed that controls provided more robust correction when matched to the fluorescence phenotype of the construct being corrected, subsequent experiments matched the genetic background of controls to the promoters tested.

To summarize phenotypes estimated for each line isolated, means and SDs were calculated for independent measurements of population medians and CVs across replicated samples. Any individual replicate that was more than 4 MAD outside of other estimates for the same line was called as an outlier and excluded from further analysis. Only lines with at least 3 independent replicates passing all quality control filters were included in downstream data analysis. These stringent quality control procedures resulted in differences in the total number of lines represented across conditions for different promoter constructs (*SI Appendix*, Table S4).

Characterization of Mutational Spectra. Statistical analyses to characterize mutational spectra across promoters were performed in R (69).

To characterize asymmetry in distributions of expression levels, expression levels of lines were divided into groups with expression greater than (increases) and less than (decreases) the median sham expression level.

The observation that sham-treated lines differed in their variability among promoters (*SI Appendix*, Fig. S1*A*) led us to calculate a *Z* score as a metric for capturing the increase in variability due to EMS treatment alone across promoters. To calculate mutational *Z* scores, the median of sham phenotype for each promoter was subtracted from each line's YFP expression value and the resulting quantity was divided by the SD among the sham-treated lines for that promoter. The resulting metric was centered on 0 and expressed in units representing SDs among unmutated individuals expressing a matched promoter construct (*SI Appendix*, Fig. S1*B*).

To describe differences in the shapes of these distributions of mutational effects on this *Z*-score scale, we explored a variety of metrics. Summary statistics like sample mean, variance (or SD), skewness, and kurtosis are commonly used to describe distributions; however, these measures have been shown to be particularly vulnerable to influence by outliers (41). More robust statistical measures can be used to describe distribution shape while controlling the impact of outliers in situations where sample size is limiting (39). Here, we apply the MAD to characterize distribution breadth in place of SD (70), MC to characterize distribution bias in place of skewness (71), and LMC and RMC to characterize the location of distribution tail weight in place of kurtosis. By downsampling data from a previously published mutagenesis experiment including more than 1,200 mutagenized lines, we illustrated that sample median, MAD, MC, and LMC and RMC provide more robust and repeatable characterizations of distribution shape (*SI Appendix*, Figs. S2–S5). Then, to identify the combination of variables explaining differences between mutational distributions of different promoters, we performed a principal-component analysis (72) on robust estimators of moments extracted from *Z*-score distributions across promoters (*SI Appendix*, Fig. S6).

Promoter-specific mutational distributions of *Z*-scores were visualized by generating stacked density plots using ggplot2 (73). To test for differences in the shapes of distributions of mutational effects between promoters, we applied the nonparametric AD *k*-sample test (74, 75) to identify pairwise differences between different promoter mutational distributions, using the BH procedure to control the false-discovery rate in these multiple pairwise tests at 5%.

Correlation of Promoter and Population Parameters with Mutational Summary Statistics. Promoter properties were collected from the literature (59–62, 76). Linear models predicting summary statistics MC and MAD independently were tested including all promoter property correlates as additive effects. Given the small number of genes involved, the relationship to promoter properties was explored both for continuous metrics and by dividing continuous data into categories of low and high values around the median. A process of model simplification was used to identify predictors explaining variation in MAD or MC, and a BH multiple test correction was performed. Population polymorphism quantified as variance in expression among 22 natural isolates (77) was also tested for a significant relationship with MAD and MC.

Evolutionary Simulations. To illustrate the consequences of the mutational distributions reported here for evolutionary predictions under neutrality, we simulated evolution of an asexual population of individuals randomly sampling mutations impacting the focal promoter and tracked the trajectory of the mean population phenotype over time. For each promoter, populations were initiated by sampling a starting phenotype for each individual (n = 1,000) from a smoothed version of the sham-treated population. Each generation each individual mutated with a probability determined based on the average estimate of per-generation rate of point mutations ($\sim 1.67 \times 10^{10}$ bp/generations) detected in MA studies (27) multiplied by the *S. cerevisiae* genome size (1.25×10^7 bp). When individuals mutated, they drew a mutational effect size from the distribution of expression levels estimated for EMS-treated lines and multiplied their current phenotype by that effect size. Simulations were performed using mutational effects drawn from both relative expression (Figs. 4 and 5) and *Z*-score (*SI Appendix*, Fig. S10) scales. Individuals were randomly selected for inclusion in the population each generation. Simulations ran for 50,000 generations and 500 replicate simulations were performed for each promoter. To contrast these results with more typical evolutionary Brownian motion predictions based on an assumption of normally distributed mutational effects, we also ran versions of the simulation from each promoter that were identical except that the mutational effects were drawn from a normal distribution with mean of 0 and variance based on variance among all EMS-treated lines for the promoter. For each promoter, evolutionary trajectories were described by the distribution of population means for each generation among independent replicates, and total divergence within each simulation was summarized by the distribution of population means at generation 50,000.

To ask whether differences among models and among genes altered the neutral divergence predicted, we fit a linear model predicting the median phenotype of each replicate population after 50,000 generations based on promoter identity and model identity (Log Median Expression at Generation 50 K ~ Promoter × Model Type) and use an ANOVA *F* test to assess fit of full and reduced models. To show how differences in distribution shape related to changes in evolutionary predictions under the 2 models, we summarized population mean expression among all replicates with a median and then predicted this median expression for each gene based on robust summary statistics calculated for each distribution of mutational effects (Log Grand Median Generation 50 K ~ MAD + MC + LMC + RMC). We performed this procedure separately for the Brownian and full empirical models, and performed model simplification dropping variables with no explanatory power to identify the summary statistics predicting expression divergence in each case.

Distribution Fitting. To identify the family of probability distributions best fitting the empirically defined distributions of mutational effects, maximum-likelihood estimation was performed to identify parameters and log-likelihoods for the Gaussian, symmetric LaPlace, and asymmetric LaPlace distributions given the empirical data. This procedure was performed independently for the EMS-treated populations of each promoter. BIC values were calculated for each fit to identify the best-supported model while appropriately penalizing for differences in number of parameters among the distributions (2 for Gaussian and symmetric LaPlace, 3 for asymmetric LaPlace). ΔBIC values were calculated for each fit by subtracting the BIC of the model with the lowest BIC values from all others. We took ΔBIC values greater than 10 as a signal of poor support for a given model.

Access to Data and Analysis Scripts. Raw flow cytometry data are available at http://flowrepository.org/ (records FR-FCM-ZYUW and FR-FCM-ZZNR).

Datasets are as follows: Dataset S1, primers used to generate and sequence confirm reporter constructs; Dataset S2, R code for processing raw .fcs files, normalizing phenotypes by plate controls, filtering outliers, and calculating mean phenotypes by promoter; Dataset S3, R code used to contrast mean phenotypes among promoters; and Dataset S4, R code for evolutionary simulations.

Additional data files required to reproduce this work are available from DeepBlue Data, https://deepblue.lib.umich.edu/. These include .zip files containing 1) layout spreadsheet with experimental metadata linking .fcs files to samples, and 2) processed data file including mean phenotypes.

ACKNOWLEDGMENTS. We thank current and past members of the P.J.W. laboratory for discussions informing this work, particularly Mark Hill and Brian Metzger, who provided helpful feedback on the manuscript draft. We thank Kerry Geiler-Samerotte, Max Reuter, and an anonymous reviewer for constructive feedback that improved this work. We thank the University of Michigan Center for Chemical Genomics and Flow Cytometry Core for access to flow cytometry equipment. We gratefully acknowledge the following funding sources that supported this work: NIH F32 GM115198 (to A.H.D.); EMBO ALTF 1114-2012 (to F.D.); and NSF MCB-1021398, NIH 1R35 GM118073, and NIH R01 GM108826 (to P.J.W.).

Downloaded at Palestinian Territory, occupied on December 23, 2021
www.manaraa.com

1. J. C. Fay, P. J. Wittkopp, Evaluating the role of natural selection in the evolution of gene regulation. *Heredity* **100**, 191–199 (2008).
2. A. Hodgins-Davis, J. P. Townsend, Evolving gene expression: From G to E to GxE. *Trends Ecol. Evol. (Amst.)* **24**, 649–658 (2009).
3. B. Lemos, C. D. Meiklejohn, M. Cáceres, D. L. Hartl, Rates of divergence in gene expression profiles of primates, mice, and flies: Stabilizing selection and variability among functional categories. *Evolution* **59**, 126–137 (2005).
4. D. Brawand *et al*., The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
5. A. Stoltzfus, L. Y. Yampolsky, Climbing mount probable: Mutation as a cause of nonrandomness in evolution. *J. Hered.* **100**, 637–647 (2009).
6. A. Stoltzfus, D. M. McCandlish, Mutational biases influence parallel adaptation. *Mol. Biol. Evol.* **34**, 2163–2172 (2017).
7. D. Halligan, P. Keightley, Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **40**, 151–172 (2009).
8. E. Hine, D. E. Runcie, K. McGuigan, M. W. Blows, Uneven distribution of mutational variance across the transcriptome of *Drosophila serrata* revealed by high-dimensional analysis of gene expression. *Genetics* **209**, 1319–1328 (2018).
9. C. R. Landry, B. Lemos, S. A. Rifkin, W. J. Dickinson, D. L. Hartl, Genetic properties influencing the evolvability of gene expression. *Science* **317**, 118–121 (2007).
10. D. R. Denver *et al*., The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* **37**, 544–548 (2005).
11. A. Konrad *et al*., Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7386–7391 (2018).
12. S. A. Rifkin, D. Houle, J. Kim, K. P. White, A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* **438**, 220–223 (2005).
13. K. McGuigan *et al*., The nature and extent of mutational pleiotropy in gene expression of male *Drosophila serrata*. *Genetics* **196**, 911–921 (2014).
14. W. Huang *et al*., Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *eLife* **5**, e14625 (2016).
15. M. Lynch, The rate of polygenic mutation. *Genet. Res.* **51**, 137–148 (1988).
16. M. A. White, Understanding how cis-regulatory function is encoded in DNA sequence using massively parallel reporter assays and designed sequences. *Genomics* **106**, 165–170 (2015).
17. R. P. Patwardhan *et al*., Massively parallel functional dissection of mammalian enhancers in vivo. *Nat. Biotechnol.* **30**, 265–270 (2012).
18. G. Hornung, M. Oren, N. Barkai, Nucleosome organization affects the sensitivity of gene expression to promoter mutations. *Mol. Cell* **46**, 362–368 (2012).
19. I. Mogno, J. C. Kwasnieski, B. A. Cohen, Massively parallel synthetic promoter assays reveal the in vivo effects of binding site variants. *Genome Res.* **23**, 1908–1915 (2013).
20. N. M. Belliveau *et al*., Systematic approach for dissecting the molecular mechanisms of transcriptional regulation in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4796–E4805 (2018).
21. P. J. Wittkopp, Genomic sources of regulatory variation in *cis* and in *trans*. *Cell. Mol. Life Sci.* **62**, 1779–1783 (2005).
22. F. W. Albert, J. S. Bloom, J. Siegel, L. Day, L. Kruglyak, Genetics of *trans*-regulatory variation in gene expression. *eLife* **7**, e35471 (2018).
23. P. J. Wittkopp, B. K. Haerum, A. G. Clark, Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**, 85–88 (2004).
24. B. P. H. Metzger *et al*., Contrasting frequencies and effects of *cis*- and *trans*-regulatory mutations affecting gene expression. *Mol. Biol. Evol.* **33**, 1131–1146 (2016).
25. G. Hornung *et al*., Noise-mean relationship in mutated promoters. *Genome Res.* **22**, 2409–2417 (2012).
26. A. Klungland, K. Laake, E. Hoff, E. Seeberg, Spectrum of mutations induced by methyl and ethyl methanesulfonate at the hprt locus of normal and tag expressing Chinese hamster fibroblasts. *Carcinogenesis* **16**, 1281–1285 (1995).
27. Y. O. Zhu, M. L. Siegal, D. W. Hall, D. A. Petrov, Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2310–E2318 (2014).
28. M. Lynch *et al*., A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9272–9277 (2008).
29. C. J. Maclean *et al*., Deciphering the genic basis of yeast fitness variation by simultaneous forward and reverse genetics. *Mol. Biol. Evol.* **34**, 2486–2502 (2017).
30. J. Peter *et al*., Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature* **556**, 339–344 (2018).
31. G. I. Lang, A. W. Murray, Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* **178**, 67–82 (2008).
32. J. D. Gruber, K. Vogel, G. Kalay, P. J. Wittkopp, Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: Frequency, effects, and dominance. *PLoS Genet.* **8**, e1002497 (2012).
33. F. Duveau *et al*., Mapping small effect mutations in *Saccharomyces cerevisiae*: Impacts of experimental design and mutational properties. *G3 (Bethesda)* **4**, 1205–1216 (2014).
34. A. Hodgins-Davis, P. J. Wittkopp, Empirical measures of mutational effects define neutral models of regulatory evolution in Saccharomyces cerevisiae. FlowRepository. https://flowrepository.org/id/FR-FCM-ZYUW. Deposited January 14, 2019.
35. B. P. H. Metzger, P.J. Wittkopp, Contrasting frequencies and effects of cis- and trans-regulatory mutations affecting gene expression. FlowRepository. https://flowrepository.org/id/FR-FCM-ZZNR. Deposited December 31, 2015.
36. A. Hodgins-Davis, P. J. Wittkopp, Empirical measures of mutational effects define neutral models of regulatory evolution in Saccharomyces cerevisiae. Deep Blue Data. https://doi.org/10.7302/0dvr-k169. Deposited 27 August 2019.
37. F. Duveau *et al*., Fitness effects of altering gene expression noise in *Saccharomyces cerevisiae*. *eLife* **7**, e37272 (2018).
38. D. Houle, B. Morikawa, M. Lynch, Comparing mutational variabilities. *Genetics* **143**, 1467–1483 (1996).
39. T.-H. Kim, H. White, On more robust estimation of skewness and kurtosis. *Finance Res. Lett.* **1**, 56–73 (2004).
40. G. Brys, M. Hubert, A. Struyf, Robust measures of tail weight. *Comput. Stat. Data Anal.* **50**, 733–759 (2006).
41. M. A. Medina, E. Ronchetti, Robust statistics: A selective overview and new directions. *Wiley Interdiscip. Rev. Comput. Stat.* **7**, 372–393 (2015).
42. H. A. Orr, Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* **149**, 2099–2104 (1998).
43. N. H. Barton, A. M. Etheridge, A. Véber, The infinitesimal model: Definition, derivation, and implications. *Theor. Popul. Biol.* **118**, 50–73 (2017).
44. R. Kita, S. Venkataram, Y. Zhou, H. B. Fraser, High-resolution mapping of *cis*-regulatory variation in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E10736–E10744 (2017).
45. T. F. C. Mackay, R. F. Lyman, M. S. Jackson', Effects of P element insertions on quantitative traits in *Drosophila melanogaster*. *Genetics* **130**, 315–332 (1992).
46. E. Santiago, J. Albornoz, A. Domínguez, M. A. Toro, C. López-Fanjul, The distribution of spontaneous mutations on quantitative traits and fitness in *Drosophila melanogaster*. *Genetics* **132**, 771–781 (1992).
47. M. A. López, C. López-Fanjul, Spontaneous mutation for a quantitative trait in *Drosophila melanogaster*. II. Distribution of mutant effects on the trait and fitness. *Genet. Res.* **61**, 117–126 (1993).
48. R. F. Lyman, F. Lawrence, S. V. Nuzhdin, T. F. C. Mackay, Effects of single P-element insertions on bristle number and viability in *Drosophila melanogaster*. *Genetics* **143**, 277–292 (1996).
49. P. D. Keightley, O. Ohnishi, P. D. Keightley, EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. *Genetics* **148**, 753–766 (1998).
50. J. D. Fry, P. D. Keightley, S. L. Heinsohn, S. V. Nuzhdin, New estimates of the rates and effects of mildly deleterious mutation in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 574–579 (1999).
51. J. D. Fry, S. L. Heinsohn, Environment dependence of mutational parameters for viability in *Drosophila melanogaster*. *Genetics* **161**, 1155–1167 (2002).
52. T. F. C. Mackay, R. F. Lyman, F. Lawrence, Polygenic mutation in *Drosophila melanogaster*: Mapping spontaneous mutations affecting sensory bristle number. *Genetics* **170**, 1723–1735 (2005).
53. B. Park *et al*., Distributions of mutational effects and the estimation of directional selection in divergent lineages of *Arabidopsis thaliana*. *Genetics* **206**, 2105–2117 (2017).
54. X. -S. Zhang, W. G. Hill, The anomalous effects of biased mutation revisited: Mean–optimum deviation and apparent directional selection under stabilizing selection. *Genetics* **179**, 1135–1141 (2008).
55. J. J. Welch, D. Waxman, Nonequivalent loci and the distribution of mutant effects. *Genetics* **161**, 897–904 (2002).
56. J. G. Schraiber, M. J. Landis, Sensitivity of quantitative traits to mutational effects and number of loci. *Theor. Popul. Biol.* **102**, 85–93 (2015).
57. B. Munsky, G. Li, Z. R. Fox, D. P. Shepherd, G. Neuert, Distribution shapes govern the discovery of predictive models for gene regulation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7533–7538 (2018).
58. A. Hodgins-Davis, D. P. Rice, J. P. Townsend, Gene expression evolves under a house-of-cards model of stabilizing selection. *Mol. Biol. Evol.* **32**, 2130–2140 (2015).
59. J. R. S. Newman *et al*., Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).
60. A. D. Basehoar, S. J. Zanton, B. F. Pugh, Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699–709 (2004).
61. I. Tirosh, N. Barkai, Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* **18**, 1084–1091 (2008).
62. L. M. Steinmetz *et al*., Systematic screen for human disease genes in yeast. *Nat. Genet.* **31**, 400–404 (2002).
63. F. C. Holstege *et al*., Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
64. A. M. Deutschbauer, R. W. Davis, Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* **37**, 1333–1340 (2005).
65. L. N. Dimitrov, R. B. Brem, L. Kruglyak, D. E. Gottschling, Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of *Saccharomyces cerevisiae* S288C strains. *Genetics* **183**, 365–383 (2009).
66. F. Hahne *et al*., flowCore: A bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* **10**, 106 (2009).
67. K. Lo, F. Hahne, R. R. Brinkman, R. Gottardo, flowClust: A bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics* **10**, 145 (2009).
68. L. Wang, A. K. Gaigalas, Development of multicolor flow cytometry calibration standards: Assignment of equivalent reference fluorophores (ERF) unit. *J. Res. Natl. Inst. Stand. Technol.* **116**, 671–683 (2011).
69. R Core Team, R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, 2017). https://www.r-project.org/. Accessed 20 September 2019.
70. D. Meyer *et al*., R Package "e1071" (Misc Functions of the Department of Statistics, Probability Theory Group) (Vienna University of Technology, Vienna, 2018). https://CRAN.R-project.org/package=e1071. Accessed 25 October 2018.
71. Maechler M *et al*. Robustbase: Basic Robust Statistics R Package (2017). https://cran.r-project.org/web/packages/robustbase/index.html. Accessed 25 October 2018.
72. S. Lê, J. Josse, F. Husson, FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).
73. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009).
74. S. Engmann, D. Cousineau, Comparing distributions: The two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. *J. Appl. Quant. Methods* **6**, 1–17 (2011).
75. F. Scholtz, A. Zhu, R Package kSamples: K-Sample Rank Tests and Their Combinations (2018).
76. J. M. Cherry *et al*., *Saccharomyces* genome database: The genomics resource of budding yeast. *Nucleic Acids Res.* **40**, D700–D705 (2012).
77. D. A. Skelly *et al*., Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res.* **23**, 1496–1504 (2013).
78. F. Duveau, W. Toubiana, P. J. Wittkopp, Fitness effects of *cis*-regulatory variants in the *Saccharomyces cerevisiae* TDH3 promoter. *Mol. Biol. Evol.* **34**, 2908–2912 (2017).

Downloaded at Palestinian Territory, occupied on December 23, 2021
www.manaraa.com