

Testing the neutral hypothesis of phenotypic evolution

Wei-Chin Ho¹, Yoshikazu Ohya², and Jianzhi Zhang^{1,*}

¹Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

²Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba Prefecture 277-8562, JAPAN

*Correspondence to:

Jianzhi Zhang

Department of Ecology and Evolutionary Biology

University of Michigan

1075 Natural Science Building

830 North University Avenue

Ann Arbor, MI 48109

Phone: 734-763-0527

Fax: 734-763-0544

Email: jianzhi@umich.edu

It is generally accepted that a large fraction of genomic sequence variations within and between species are neutral or nearly so¹. Whether the same is true for phenotypic variations is a central question in biology²⁻⁷. On the one hand, numerous phenotypic adaptations have been documented^{2,8,9} and even Kimura, the champion of the neutral theory of molecular evolution, believed in widespread adaptive phenotypic evolution¹. On the other hand, phenotypic studies are strongly biased toward traits that are likely to be adaptive⁹, contrasting genomic studies that are typically unbiased. It is thus desirable to test the neutral hypothesis of phenotypic evolution using traits irrespective of their potential involvement in adaptation. Here we present such a test for 210 morphological traits measured in multiple strains of the yeast *Saccharomyces cerevisiae* and two related species. Our test is based on the premise that, under neutrality, the rate of phenotypic evolution declines as the trait becomes more important to fitness, analogous to the neutral paradigm that functional genes evolve more slowly than functionless pseudogenes¹⁰.

Neutrality is rejected in favor of adaptation if important traits evolve faster than less important ones, parallel to the demonstration of molecular adaptation when a functional gene evolves faster than pseudogenes. After controlling for the mutational size, we find faster evolution of more important morphological traits within and between species. By contrast, an analysis of 3466 yeast gene expression traits fails to reject neutrality. Thus, yeast morphological evolution is largely adaptive, but the same may not apply to other classes of phenotypes.

Analogous to the neutral hypothesis of molecular evolution¹, the neutral hypothesis of phenotypic evolution allows the presence of purifying selection; neutrality is rejected only when positive selection is invoked. Under the neutral hypothesis, compared with traits that are

relatively unimportant to fitness, relatively important traits should be subject to stronger purifying selection and evolve more slowly given the same speed of mutational input (**Fig. 1a**). However, if relatively important traits evolve faster than relatively unimportant traits, the neutral hypothesis would no longer hold and the only reasonable explanation would be stronger positive selection acting on relatively important traits than relatively unimportant ones (**Fig. 1b**). This test of phenotypic neutrality differs from previous tests^{3,4,11-15}, which consider only one trait at a time and effectively require the intensity of positive selection to surpass that of purifying selection to reject neutrality. Because this requirement is sufficient but not necessary for demonstrating positive selection, it is replaced with the criterion of a positive correlation between trait importance and evolutionary rate to improve the power of the test (see Methods and Discussion).

We first tested the neutral hypothesis of phenotypic evolution in a set of 210 morphological traits chosen purely based on the feasibility of measurement¹⁶ rather than potential roles in adaptation (**Supplementary Data 1**). These traits were quantified by analyzing fluorescent microscopic images of triple-stained cells¹⁶ from 37 natural strains of *S. cerevisiae*¹⁷. For a given trait, we defined the phenotypic difference between two strains as the absolute difference in their trait value, relative to their average trait value. Phenotypic differences were corrected for potential environmental heterogeneity in the measurement and sampling error to allow among-trait comparison (**Supplementary Data 2**; see Methods). We then estimated, for each trait, the mean evolutionary distance (*ED*) among all 666 pairs of the 37 strains by averaging their corrected pairwise phenotypic differences.

To test the neutral hypothesis, we used measures of trait importance (*TI*) for the 210 traits (**Supplementary Data 1**), where *TI* is 100 times the fitness effect caused by 1% change in trait

value, estimated using the fitness and phenotype data of thousands of single gene deletion strains of *S. cerevisiae*¹⁸. We found that mean *ED* decreases with *TI* (**Fig. 2a**), meaning that relatively important traits evolve more slowly than relatively unimportant ones. However, the rate of phenotypic evolution is determined by both the rate of mutational input and the direction and magnitude of natural selection. The rate of mutational input for a trait is the average effect size of a random mutation on the trait (i.e., mutational size or *MS*) multiplied by the mutation rate per genome per generation. Because the mutation rate is the same for all traits, we need only consider *MS*. We estimated the *MS* for a trait by the mean phenotypic effect of 4718 individual gene deletions on the trait¹⁸ (**Supplementary Data 1**). Although spontaneous mutations are expected to have smaller phenotypic effects than gene deletions, it is reasonable to assume that the average phenotypic effect of spontaneous mutations on a trait is approximately proportional to that of gene deletions on the trait. In other words, *MS* can be used as a proxy of spontaneous mutational size when different traits are compared. As was previously discovered¹⁸, *MS* decreases precipitously with *TI* (**Fig. 2a**). This observation indicates that relatively important traits are affected by mutations to a smaller degree than are relatively unimportant traits, which has likely resulted from stronger selection for mutational robustness of more important traits¹⁸. To control the impact of mutational input on the rate of phenotypic evolution, we measured the evolutionary rate of a trait in the unit of its mutational size by dividing mean *ED* by *MS* for each trait. We found mean *ED/MS* to increase significantly with *TI* (**Fig. 2b**), suggesting that relatively important traits evolve faster than relatively unimportant traits in the unit of mutational size. This finding is inconsistent with the neutral hypothesis of phenotypic evolution (**Fig. 1a**), but supports the adaptive hypothesis (**Fig. 1b**).

To exclude the possibility that the above result is an artifact of our statistical analysis, we

compared the wild-type BY strain with 666 randomly picked gene deletion strains of the BY background, under the premise that phenotypic differences between deletion strains and the wild-type should not be adaptive. As expected, there is no significant correlation between mean *ED/MS* and *TI* (**Fig. 2b**), where mean *ED* for a trait is the average phenotypic difference between these deletion strains and the wild-type for the trait.

Because the above *MS* and *TI* were estimated in haploid yeasts, while the 37 natural strains studied are diploid, we also estimated *MS* and *TI* using recently published morphological data of 130 diploid gene deletion strains¹⁹. We confirmed that the significant positive correlation between *ED/MS* and *TI* holds (**Extended Data Fig. 1**).

To examine whether the detected adaptive signal among the 37 natural strains is attributable to a small number of strains or is a general phenomenon of the species, we estimated the rank correlation (ρ) between *ED/MS* and *TI* for each of the 666 strain pairs, using the *ED* value of the strain pair. We found ρ to be positive for the vast majority of the strain pairs (**Fig. 2c**), suggesting pervasive adaptive morphological evolution in *S. cerevisiae*.

Some of the 210 morphological traits are genetically highly correlated^{18,20}. To exclude the possibility that the adaptive signal is an artifact of the use of correlated traits, we estimated 210 principal component traits from the original traits (see Methods). Analysis of the principal component traits, which are independent from one another, shows even stronger signals of adaptive evolution (**Extended Data Fig. 2**).

If the prevalent positive ρ values among the 37 natural strains (**Fig. 2c**) truly arise from positive selection, we should expect that (i) the overall rate of morphological evolution is greater for strain pairs with higher ρ values and (ii) this rate disparity is primarily reflected in relatively

important traits rather than relatively unimportant ones. We defined the overall rate of morphological evolution between two strains by their morphological dissimilarity across all 210 traits divided by their fractional genomic sequence difference. Consistent with our expectation, the rate of morphological evolution increases with ρ ($P = 0.007$, one-tail partial Mantel test with phylogenetic permutation to correct for both the nonindependence among strain pairs and phylogenetic relationships in the data; see Methods). Additionally, when we separated the 210 traits into two equal-sized bins based on TI , the correlation between the rate of morphological evolution and ρ remained significantly positive for the 105 traits with relatively high TI ($P < 0.001$) but not for the 105 traits with relatively low TI ($P = 0.805$). These observations support that, the greater the ρ value relative to 0, the stronger the positive selection on the morphological traits, especially relatively important ones.

What factors determine the ρ value of a strain pair? Because the 210 morphological traits were chosen purely based on experimental feasibility, we do not expect the detected adaptive signals to correlate with any obvious genetic or ecological factor. Indeed, using the partial Mantel test, we found no significant correlation between the ρ value of two strains and the strains' difference in genome sequence, ecological environment, population membership, or geographic location (**Extended Data Table 1**). These findings are consistent with a previous analysis showing that the morphological similarities among the 37 natural strains cannot be explained by the strains' similarities in population history or ecological environment¹⁷. Hence, the selective agents behind the detected morphological adaptations are unclear.

To test the neutral hypothesis of morphological evolution beyond the species level, we collected comparable morphological data from two strains of *S. paradoxus*, the sister species of *S. cerevisiae*, and one strain of their outgroup species *S. mikatae* (see Methods). We first calculated

the mean ED/MS between the 37 *S. cerevisiae* strains and *S. paradoxus* strain N17 for each trait (**Supplementary Data 3**). Across the 210 traits, we observed a significantly positive correlation between mean ED/MS and TI (**Fig. 3a**). A similar pattern was observed between *S. cerevisiae* and *S. paradoxus* strain IFO1804 (**Fig. 3b**). Between *S. cerevisiae* and the more distantly related species of *S. mikatae* (**Supplementary Data 3**), however, the positive correlation is no longer significant (**Fig. 3c**), probably because of the lack of prevalent positive selection or a reduced statistical power as a result of using MS and TI values estimated from *S. cerevisiae* in tests involving distantly related species.

To investigate the generality of the above findings of adaptive phenotypic evolution, we turned to another class of traits that can be chosen regardless of their potential roles in adaptation: 3466 gene expression traits, each being the mRNA expression level of a yeast gene in a rich medium. Using microarray gene expression data, we quantified ED between two *S. cerevisiae* strains for each trait (**Supplementary Data 4**). We used two approaches to estimate MS on each trait. First, we estimated it from the microarray expression data of 1486 gene deletion strains, as in the case of morphological traits¹⁸. Second, we estimated it by the standard deviation (SD_m) in relative expression level among four mutation accumulation lines²¹, arguably a more natural measure of mutational size. TI of the expression level of a gene was measured by the fitness reduction caused by the deletion of the gene¹⁸ (see Methods). Similar to the morphological data, gene expression data showed a negative correlation between ED and TI and a negative correlation between MS (or SD_m) and TI (**Table 1**). However, contrary to the morphological traits, expression traits exhibited a significantly negative correlation between ED/MS (or ED/SD_m) and TI (**Table 1**). While adaptive evolution of the expression levels of some yeast genes have been suggested²²⁻²⁴, overall we found relatively important genes to evolve more slowly in

expression level than relatively unimportant ones, consistent with the neutral hypothesis. The similar results obtained from the use of MS and SD_m support that MS is a suitable proxy for mutational size. Here we are restricted to intra-specific analysis, because the use of different probes in the microarrays of different species prohibits a reliable comparison of the rate of expression evolution among genes. Use of other types of expression data such as mRNA sequencing data is currently infeasible, because no comparable data are available for estimating MS .

In summary, our analysis of 210 yeast morphological traits with no *a priori* bias toward adaptive evolution reveals strong signals of adaptive evolution. To the best of our knowledge, this is the first demonstration of widespread adaptive evolution of a large set of random phenotypic traits in any organism. If these 210 traits are representative of yeast morphological traits, we must conclude that morphological evolution in yeast is frequently adaptive. Whether the same conclusion could be drawn in other species is unknown, but the methodology developed here is generally applicable, especially to genetic model organisms, in which mutational size estimation and trait importance estimation are relatively straightforward. Although it is essential to use a random set of traits to evaluate the prevalence of adaptive phenotypic evolution, using such traits means that the selective agent would be difficult to discern if adaptation is detected, as in the present case. Future work should attempt to identify the selective agents acting on these morphological traits, which are necessary for a complete understanding of phenotypic adaptation.

It should be emphasized that, although the bar for rejecting neutrality is likely lower in our test than in most previous tests^{3,4,11-13}, it remains quite high. This is because important traits that could be subject to strong positive selection are also expected to be under strong purifying

selection such that an adaptive signal becomes detectable only when the difference in the strength of positive selection among traits surpasses that in the strength of purifying selection. The high bar renders claims of adaptation conservative. But, it also means that failure to reject neutrality, as in the case of yeast gene expression evolution, neither proves neutrality nor refutes adaptation. Nonetheless, based on additional experiments and tests, we recently found unambiguous evidence that the vast majority of gene expression level variations within and between yeast species are not adaptive but neutral²⁵. Together, these analyses of yeast morphological and gene expression data raise the intriguing possibility that some classes of phenotypic traits evolve generally adaptively while others neutrally.

Acknowledgements

We thank Seiko Morinaga for technical assistance and Xiaoshu Chen, Brian Metzger, and Jian-Rong Yang for valuable comments. This work was supported by U.S. National Institutes of Health research grant R01GM103232 to J.Z. and the Grants-in-Aid for Scientific Research grant 24370002 from the Ministry of Education, Culture, Sports, Science and Technology in Japan to Y.O.

Figure legends

Figure 1. Comparing evolutionary rates among traits to test the neutral hypothesis of phenotypic evolution. **a**, Under the neutral hypothesis, relatively important traits evolve more slowly than relative unimportant traits. **b**, Higher evolutionary rate for more important traits rejects the neutral hypothesis and supports the adaptive hypothesis. Each circle represents a trait.

Figure 2. Prevalent adaptive evolution of morphological traits in the yeast *Saccharomyces cerevisiae*. **a**, Mean evolution distance (ED) of 666 pairs of natural strains for a trait and the mutational size (MS) of the trait both decrease with trait importance (TI). Each dot represents a morphological trait. ρ , Spearman's rank correlation coefficient. **b**, Mean ED among 666 natural strain pairs for a trait relative to its MS increases significantly with TI , while the mean ED between 666 gene deletion strains and the wild-type relative to MS does not increase significantly with TI . **c**, Nominal p -values for the Spearman's correlation between ED/MS and TI for all 666 pairs of natural strains. The horizontal colored bars above the strain names show the ecological environments of the strains. Be, beer; Ba, bakery.

Figure 3. Adaptive morphological evolution between *Saccharomyces* species. **a**, The mean ED/MS between 37 natural *S. cerevisiae* strains and *S. paradoxus* strain N17 for a trait increases significantly with trait importance (TI). **b**, The mean ED/MS between 37 natural *S. cerevisiae* strains and the *S. paradoxus* strain IFO1804 increases significantly with TI . **c**, The mean ED/MS between 37 natural *S. cerevisiae* strains and *S. mikatae* increases (but not significantly) with TI . Each dot represents a morphological trait.

Extended Data Figure 1. Mean ED among 666 natural strain pairs of *S. cerevisiae* for a trait relative to its MS increases significantly with TI , when MS and TI are both estimated using diploid deletion strains.

Extended Data Figure 2. Prevalent adaptive evolution of morphological principal component traits in the yeast *S. cerevisiae*. **a**, Mean evolution distance (ED) of 666 pairs of natural strains for a trait and the mutational size (MS) of the trait both decrease with trait importance (TI). Each dot represents a morphological principal component trait. ρ , Spearman's rank correlation coefficient. **b**, Mean ED among 666 natural strain pairs relative to MS increases significantly with TI , while the mean ED between 666 gene deletion strains and the wild-type relative to MS does not increase significantly with TI . **c**, Nominal p -values for the Spearman's correlation between ED/MS and TI for all 666 pairs of natural strains.

METHODS

A comparison between existing tests of the neutral hypothesis of phenotypic evolution and the newly proposed test

A number of neutrality tests of phenotypic evolution exist in the literature. They all test the neutrality of one trait and can be divided into two categories based on the rationale of the test. The first category compares the observed phenotypic distance in a trait between populations or species with the neutral expectation. Positive (or purifying) selection is inferred when the observed distance is significantly larger (or smaller) than the neutral expectation. The neutral expectation has been derived by considering the stochastic process of phenotypic evolution³ or genetic models of quantitative traits^{4,13}. Various test statistics have been proposed, including for example Lande's N_e ³, Lande's F ¹⁴, Chakraborty and Nei's B_t/V_t ¹³, Turelli et al.'s MDE test¹², and Lynch's Δ ¹⁵. These tests usually require the information of effective population size and divergence time. They also require the information on the rate of mutational input such as narrow-sense heritability and mutational variance. If these parameters are unavailable, one may use an alternative approach by comparing Q_{ST} of a trait with F_{ST} of neutral loci, which are expected to be the same if the trait evolves neutrally¹¹. Positive selection is undetectable by any of the above tests unless the intensity of positive selection exceeds that of purifying selection. Because purifying selection is expected to be pervasive even for traits subject to positive selection, all tests in this category have a low power in detecting positive selection.

The second category is the QTL sign test, which relies on the information from QTL mapping of a trait. Typically, positive selection is inferred when the number of positive-effect QTLs differs significantly from that of negative-effect QTLs²⁶. Although this category of test is model-free, it cannot distinguish between positive selection and relaxation from purifying

selection without other information²³.

Our test of the neutral hypothesis of phenotypic evolution compares the evolutionary rates of many traits with different levels of importance to fitness. Under the neutral hypothesis allowing purifying selection, relatively important traits evolve more slowly than relatively unimportant traits. If relatively important traits are found to evolve more rapidly than relatively unimportant ones, neutrality is rejected in favor of positive selection. Our test is expected to be more powerful than the above first category of tests, because detecting positive selection no longer requires the intensity of positive selection to exceed that of purifying selection. Rather, the criterion is that the difference in intensity between positive and negative selection is more positive for more important traits. Our test does not suffer from the problem in the above second category of tests, because it is extremely improbable for the relative importance of a large number of traits to be reversed in a second population, compared with that in the original population where trait importance is measured.

The yeast morphological data from natural strains and gene deletion strains, gene expression data, and the fitness data of gene deletion strains used for estimating trait importance were all collected under a laboratory rich medium. While this medium does not equal the many natural environments of various yeast strains, this discrepancy does not affect our neutrality test, because all required by our test is that the same condition is used in phenotyping a trait and in estimating its trait importance. The main problem with mismatches between the experimental condition and the natural environments is that the selective agent becomes more difficult to discern.

Yeast morphological data

The morphological traits analyzed here were previously defined¹⁶. Briefly, yeast cultures were grown to 1×10^7 cells/ml in YPD or synthetic complete media. Cells were fixed with 3.7% formaldehyde and stained with fluorescein isothiocyanate-Con A, rhodaminephalloidin, and 4',6-diamidino-2-phenylindole, which simultaneously mark cell wall, actin cytoskeleton, and nuclear DNA, respectively. After digital images were acquired, cell images were collected and processed by *CalMorph*. In total, 501 morphological traits were measured. Among these morphological traits, we focused on 210 traits, which are defined for individual cells rather than cell populations, have positive trait importance values¹⁸, and were measured in all the strains analyzed here. The morphological data of *S. paradoxus* strain N17, *S. paradoxus* strain IFO 1804, and *S. mikatae* strain IFO1815 were generated with 15, 10, and five biological replicates, respectively. On average, ~60 cells were measured for each trait in each replicate. The morphological data from 37 natural strains of *S. cerevisiae* were previously collected, with five biological replicates per strain¹⁷. The morphological data of 4718 haploid and 130 diploid single gene deletion strains of *S. cerevisiae* were previously published^{16,19}.

Evolutionary distance between two strains for a morphological trait

We estimated the mean phenotypic value for a trait in a strain by first calculating the mean trait value among all cells in a replicate population and then averaging this number across all replicate populations. Let x_i and x_j be the mean phenotypic values of a trait in strains i and j , respectively. We estimated the raw evolutionary distance for the trait between strains i and j by $ED_{ij} = |x_i - x_j| / [(x_i + x_j)/2]$. Different traits have different levels of environmental variation in phenotyping, different levels of random measurement error, and different levels of among-individual stochastic phenotypic variation. To allow a comparison among traits, we corrected the

above estimated raw ED values for these factors¹⁸. Let us assume that strain i has m replicate populations, with population sizes of a_1, a_2, \dots, a_m , respectively, and strain j has n replicate populations, with population sizes of b_1, b_2, \dots, b_n , respectively. We generated 100 sets of bootstrap samples for both strain i (i.e., each having m populations with sizes of a_1, a_2, \dots, a_m) and strain j (i.e., each having n populations with sizes of b_1, b_2, \dots, b_n) using the data from the n populations of strain j . Note that the hierarchical structure of the data is retained in bootstrapping. We then estimated the average ED_{ij} across these 100 sets of bootstrapped samples, and denoted it by pseudo $ED_{j \rightarrow i}$. We similarly estimated pseudo $ED_{i \rightarrow j}$ using the data from the m populations of strain i . We then averaged $ED_{j \rightarrow i}$ and $ED_{i \rightarrow j}$ and subtracted this value from raw ED_{ij} to obtain the corrected ED_{ij} . If the corrected ED_{ij} is negative, we set it at 0. All ED values presented in the main text and figures are corrected ED_{ij} .

Mutational size

The mutational size for a trait was calculated by the average of the previously published net effect sizes of 4718 haploid single gene deletions on the trait¹⁸. Briefly, we first calculated the raw effect size (ES_{ij}) of deleting gene i on trait j as $(x_{ij} - w_j)/w_j$, where x_{ij} is the mean phenotypic value of trait j in the deletion strain i , and w_j is the corresponding value in the wild-type (averaged across 123 replicate populations). Then we generated 1000 pseudo phenotypic datasets to estimate the net $|ES|$ of gene deletion on a trait. To generate a pseudo dataset, we randomly chose one of the 123 wild-type replicate populations and picked (with replacement) from this population the same number of cells as in the actual gene-deletion data. We then calculated mean pseudo $|ES|$ across all pseudo datasets, and net $|ES|$ equals raw $|ES|$ minus mean pseudo $|ES|$ if raw $|ES| >$ mean pseudo $|ES|$ or zero if raw $|ES| <$ mean pseudo $|ES|$.

We also estimated mutational size by the same method but using the recently published morphological data of 130 diploid single gene deletion strains¹⁹ along with the five replicate populations of the corresponding wild-type (BY4743)¹⁷. Due to the low number of wild-type replicates, 100 pseudo datasets were generated.

Morphological trait importance

The trait importance (*TI*) of each of the 210 morphological traits was previously estimated from the negative slope of the linear regression between the corrected phenotypic effect of a gene deletion on the trait and the fitness of the gene deletion strain across 2779 haploid single gene deletion strains¹⁸. The deletion strains with fitness larger than 1 were not used. Briefly, for each trait *j*, we performed a linear regression $F_i = a_j - b_j(\text{net } |ES_{ij}|)$, where net $|ES_{ij}|$ is the absolute value of the net effect size of deleting gene *i* on trait *j*, and F_i is the fitness of the strain lacking gene *i* relative to the wild-type in YPD. In this regression, the estimated slope $b_j > 0$ is 100 times the reduction in fitness caused by 1% change in the phenotypic value of trait *j*, while the estimated intercept a_j is the expected fitness when net $|ES_{ij}| = 0$. Thus, b_j is a measure of the relative importance of trait *j* to fitness, or trait importance (*TI*).

Because *TI* is estimated by the correlation between phenotypic changes and fitness changes, it may not accurately reflect the causal relationship between the variation of a trait and fitness. The fact that our use of the inaccurate *TI* estimates still yields significant evidence for adaptive morphological evolution suggests that the true signal is even stronger. In other words, our results are likely to be conservative.

For diploid single gene deletion strains¹⁹, we applied the same method to calculate *TI*, but in this case only 99 strains with fitness ≤ 1 could be used.

Principal component analysis

To examine if the non-independence among the 210 morphological traits affects our results, we followed previous studies^{18,20} to perform a principal component analysis to transform the net $|ES|$ matrix \mathbf{M} (4718 genes \times 210 traits) described previously¹⁸. Note that Net $|ES|$ is the corrected absolute effect size of a gene deletion on a trait. After this function returned a coefficient matrix \mathbf{C} (210 \times 210), we calculated $\mathbf{v}' = \mathbf{v} \mathbf{C}$ (1 \times 210 principal traits) for each corrected ED vector \mathbf{v} (1 \times 210 traits) between two strains. The absolute values of \mathbf{v}' provided the corrected ED for each of the 210 orthogonal principal component traits. We also used the transformed net $|ES|$ matrix, $\mathbf{M}'=\mathbf{MC}$ (4718 genes \times 210 principal traits), to estimate trait importance for the 210 principal component traits following the method previously used¹⁸.

Mantel's test

We used Mantel's test^{27,28} to evaluate the significance of the correlation between a biological distance (e.g., morphological dissimilarity between two strains) and the correlation (ρ) between ED/MS and TI across 666 pairs of the 37 natural strains of *S. cerevisiae*. In general, Mantel's test evaluates whether the correlation between two distance matrices is significant by comparing the test statistic (z) calculated by the observed matrices and the null distribution of z calculated by randomly shuffling one of the matrices. The test statistic could be an element-by-element product or a Pearson correlation coefficient between all elements. During each shuffling, two columns are randomly picked and swapped, and the corresponding rows are also swapped. In this study, we used the correlation coefficient between all elements of two matrices

as the test statistic and generated the null distribution of the test statistic by randomly shuffling one of the matrices 1000 times.

When the dataset has underlying phylogenetic relationships such as the present dataset in which all strains are connected by a phylogeny, the significance level may be inflated because of the shared evolutionary history among lineages²⁹. To control the phylogenetic non-independence, we performed a partial Mantel test, in which the phylogenetic distance matrix is controlled³⁰. Among four different ways to implement the partial Mantel test, we used the so-called method 2, in which the residual matrix instead of the original matrix is shuffled, as suggested³¹. In addition, we used phylogenetic permutation (PP) in partial Mantel test to obtain non-inflated type-I errors³². Specifically, applying PP means that the chance to pick a strain pair to swap is proportional to their phylogenetic distance³³. The phylogenetic tree of the 37 natural strains of *S. cerevisiae* was reconstructed using the single nucleotide polymorphism (SNP) data from Maclean et al.³⁴ by the neighboring-joining method³⁵ with *p*-distance³⁶ implemented in MEGA 5.2.2³⁷, and the matrix of pairwise phylogenetic distances was obtained by APE in R³⁸.

We used the Spearman's correlation coefficient (ρ) between *ED/ES* and *TI* of a pair of strains as a measure of adaptation. By partial Mantel's test, we examined whether ρ is correlated with the following five parameters for the strain pair: morphological evolutionary rate, dissimilarity in genome sequence, dissimilarity in ecological environment, dissimilarity in population membership, and dissimilarity in geographic location. The morphological evolutionary rate between two strains was estimated as follows. First, for each trait, we ranked the 37 strains by their mean phenotypic values for the trait. Second, we calculated Spearman's correlation between the 210 ranks of one strain with those of the other strain. The rate of morphological evolution was calculated by the negative of this Spearman's correlation divided

by genomic sequence dissimilarity (see below). Genomic sequence dissimilarity was calculated by the SNP density between two strains³⁴. Ecological environment dissimilarity was set as 0 if two strains were sampled from the same ecological environment and 1 if sampled from different environments³⁴. Population membership dissimilarity was set as 0 if two strains belong to the same population based on the fastSTRUCTURE analysis of the SNP data of 190 yeast strains³⁴; otherwise, it was set as 1. Strains designated as "mosaics"³⁴ were excluded from this analysis because mosaics do not represent a population. Geographic location dissimilarity was set to 0 if two strains were sampled from the same continent; otherwise, it was set to 1. Note that CLIB382 was removed from this analysis because of the lack of SNP data in Maclean et al.³⁴. Furthermore, any strain lacking information for any parameter was removed from the analysis for that particular parameter.

Gene expression data and analysis

The rich medium microarray gene expression ratio (r) between two *S. cerevisiae* strains, RM and BY, were previously measured for thousands of genes³⁹. We defined the intra-specific evolutionary distance (ED) of the expression level for a gene by $|x_{RM}-x_{BY}|/x_{BY}$, which equals $|r-1|$, because $r = x_{RM}/x_{BY}$, where x_{RM} and x_{BY} are expression levels of the gene in RM and BY, respectively.

We used two different methods to estimate mutational size for each expression trait. The first method was to use the mean effect size of gene deletion on the expression trait across 1486 deletion lines, following a previously published method¹⁸ but using a recently published large dataset⁴⁰. The second method was to use the square root of variance in the expression level of an evolved line relative to that of the ancestral line (SD_m) among four mutation accumulation lines²¹.

The trait importance (*TI*) of a gene expression trait was defined by the fitness decrease caused by deleting the gene²², and only those genes that cause a fitness reduction of zero or more when deleted were considered. After removing genes that miss any kind of data above, we obtained our final dataset with 3466 expression traits.

REFERENCES

- 1 Kimura, M. *The Neutral Theory of Molecular Evolution*. (Cambridge University Press, 1983).
- 2 Darwin, C. *On the Origin of Species by Means of Natural Selection*. (J. Murray, 1859).
- 3 Lande, R. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314-334 (1976).
- 4 Lynch, M. & Hill, W. G. Phenotypic evolution by neutral mutation. *Evolution* **40**, 915-935 (1986).
- 5 Mayr, E. *What Evolution is*. (Basic Books, 2001).
- 6 Nei, M. The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci U S A* **104**, 12235-12242 (2007).
- 7 Futuyma, D. J. *Evolution*. 3rd edn, (Sinauer Associates, 2013).
- 8 Endler, J. A. *Natural Selection in the Wild*. (Princeton University Press, 1986).
- 9 Kingsolver, J. G. *et al*. The strength of phenotypic selection in natural populations. *Am Nat* **157**, 245-261 (2001).
- 10 Li, W. H., Gojobori, T. & Nei, M. Pseudogenes as a paradigm of neutral evolution. *Nature* **292**, 237-239 (1981).
- 11 Spitz, K. Population structure in *Daphnia obtusa* - quantitative genetic and allozymic variation. *Genetics* **135**, 367-374 (1993).
- 12 Turelli, M., Gillespie, J. H. & Lande, R. Rate tests for selection on quantitative characters during macroevolution and microevolution. *Evolution* **42**, 1085-1089 (1988).
- 13 Chakraborty, R. & Nei, M. Genetic differentiation of quantitative characters between populations or species .1. mutation and random genetic drift. *Genet Res* **39**, 303-314 (1982).
- 14 Lande, R. Statistical tests for natural selection on quantitative characters. *Evolution* **31**, 442-444 (1977).
- 15 Lynch, M. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Am Nat* **136**, 727-741 (1990).
- 16 Ohya, Y. *et al*. High-dimensional and large-scale phenotyping of yeast mutants. *Proc Natl Acad Sci U S A* **102**, 19015-19020 (2005).
- 17 Yvert, G. *et al*. Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC Syst Biol* **7**, 54 (2013).
- 18 Ho, W. C. & Zhang, J. The genotype-phenotype map of yeast complex traits: basic parameters and the role of natural selection. *Mol Biol Evol* **31**, 1568-1580 (2014).
- 19 Yang, M., Ohnuki, S. & Ohya, Y. Unveiling nonessential gene deletions that confer significant morphological phenotypes beyond natural yeast strains. *BMC Genomics* **15**, 932 (2014).
- 20 Wang, Z., Liao, B. Y. & Zhang, J. Genomic patterns of pleiotropy and the evolution of complexity. *Proc Natl Acad Sci U S A* **107**, 18034-18039 (2010).
- 21 Landry, C. R., Lemos, B., Rifkin, S. A., Dickinson, W. J. & Hartl, D. L. Genetic properties influencing the evolvability of gene expression. *Science* **317**, 118-121 (2007).
- 22 Qian, W., Ma, D., Xiao, C., Wang, Z. & Zhang, J. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep* **2**, 1399-1410 (2012).
- 23 Fraser, H. B., Moses, A. M. & Schadt, E. E. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proc Natl Acad Sci U S A* **107**, 2977-2982 (2010).

- 24 Bullard, J. H., Mostovoy, Y., Dudoit, S. & Brem, R. B. Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc Natl Acad Sci U S A* **107**, 5058-5063 (2010).
- 25 Yang, J.-R., Maclean, C. J., Park, C., Zhao, H. & Zhang, J. Intra- and inter-specific variations of gene expression levels in yeast are largely neutral. *In submission*. (2016).
- 26 Orr, H. A. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* **149**, 2099-2104 (1998).
- 27 Mantel, N. Detection of disease clustering and a generalized regression approach. *Cancer Res* **27**, 209-& (1967).
- 28 Sokal, R. R. & Rohlf, F. J. *Biometry*. (W. H. Freeman and Company, 1995).
- 29 Felsenstein, J. Phylogenies and the comparative method. *Am Nat* **125**, 1-15 (1985).
- 30 Smouse, P. E., Long, J. C. & Sokal, R. R. Multiple-regression and correlation extensions of the Mantel test of matrix correspondence. *Syst Zool* **35**, 627-632 (1986).
- 31 Legendre, P. Comparison of permutation methods for the partial correlation and partial Mantel tests. *J Stat Comput Sim* **67**, 37-73 (2000).
- 32 Harmon, L. J. & Glor, R. E. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution* **64**, 2173-2178 (2010).
- 33 Lapointe, F. J. & Garland, T. A generalized permutation model for the analysis of cross-species data. *J Classif* **18**, 109-127 (2001).
- 34 Maclean, C. J. *et al.* Deciphering the genic basis of environmental adaptations by simultaneous forward and reverse genetics in *Saccharomyces cerevisiae*. *BioRxiv* doi: <http://dx.doi.org/10.1101/087510> (2016).
- 35 Saitou, N. & Nei, M. The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425 (1987).
- 36 Nei, M. & Kumar, S. *Molecular Evolution and Phylogenetics*. (Oxford University Press, 2000).
- 37 Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-2739 (2011).
- 38 Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289-290 (2004).
- 39 Brem, R. B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A* **102**, 1572-1577 (2005).
- 40 Kemmeren, P. *et al.* Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157**, 740-752 (2014).

Table 1. Testing the neutral hypothesis of yeast gene expression evolution

Variables correlated	Spearman's ρ	P-value
ED, TI	-0.147	3.7e-18
MS, TI	-0.142	3.9e-17
$ED/MS, TI$	-0.087	3.2e-07
SD_m, TI	-0.153	1.7e-19
$ED/SD_m, TI$	-0.060	4.5e-04

ED , evolutionary distance in gene expression level between two *S. cerevisiae* strains.

TI , gene expression trait importance.

MS , mutational size measured by gene deletion.

SD_m , standard deviation in relative gene expression level among mutation accumulation lines.

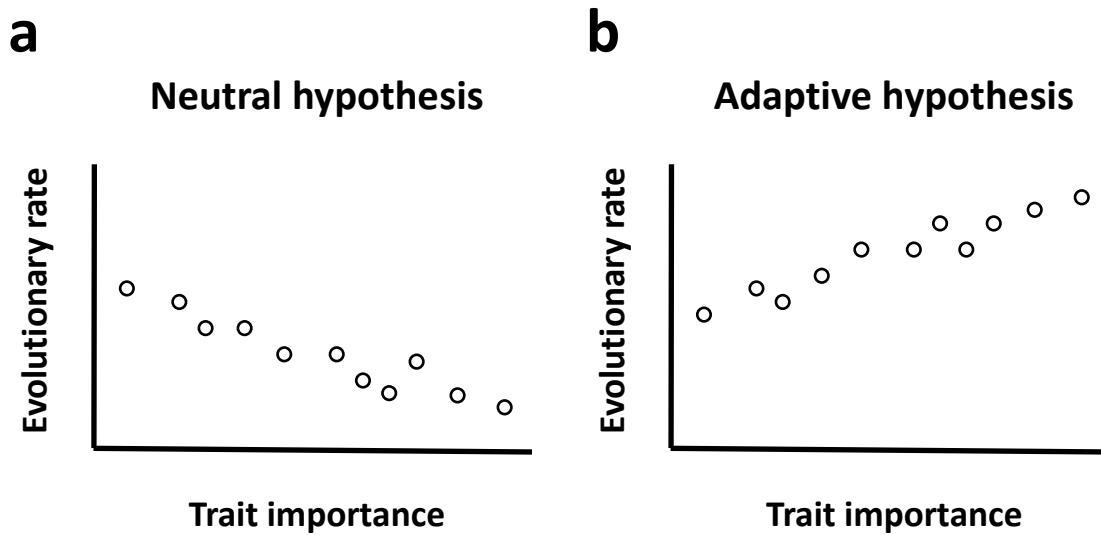


Figure 1

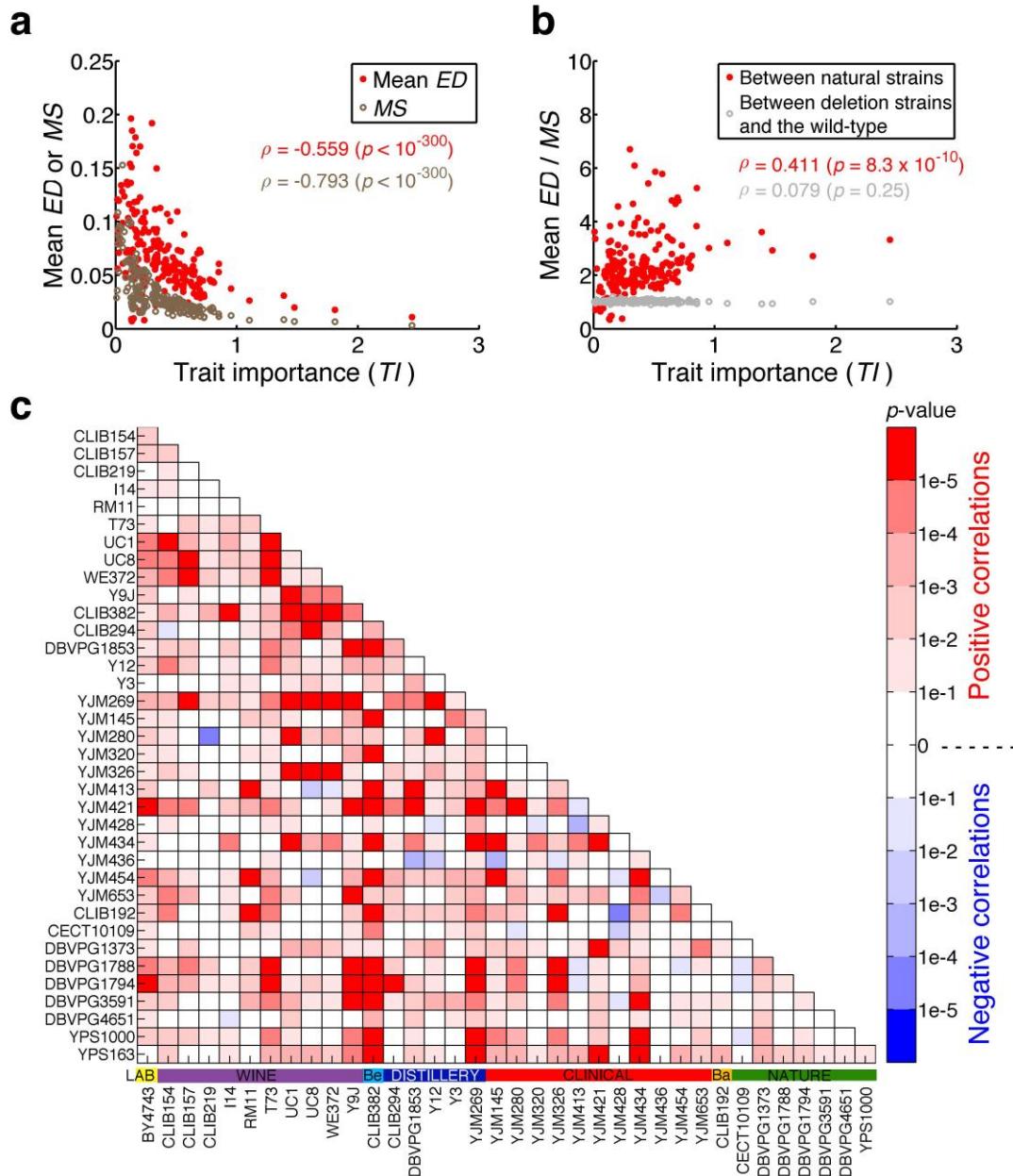


Figure 2

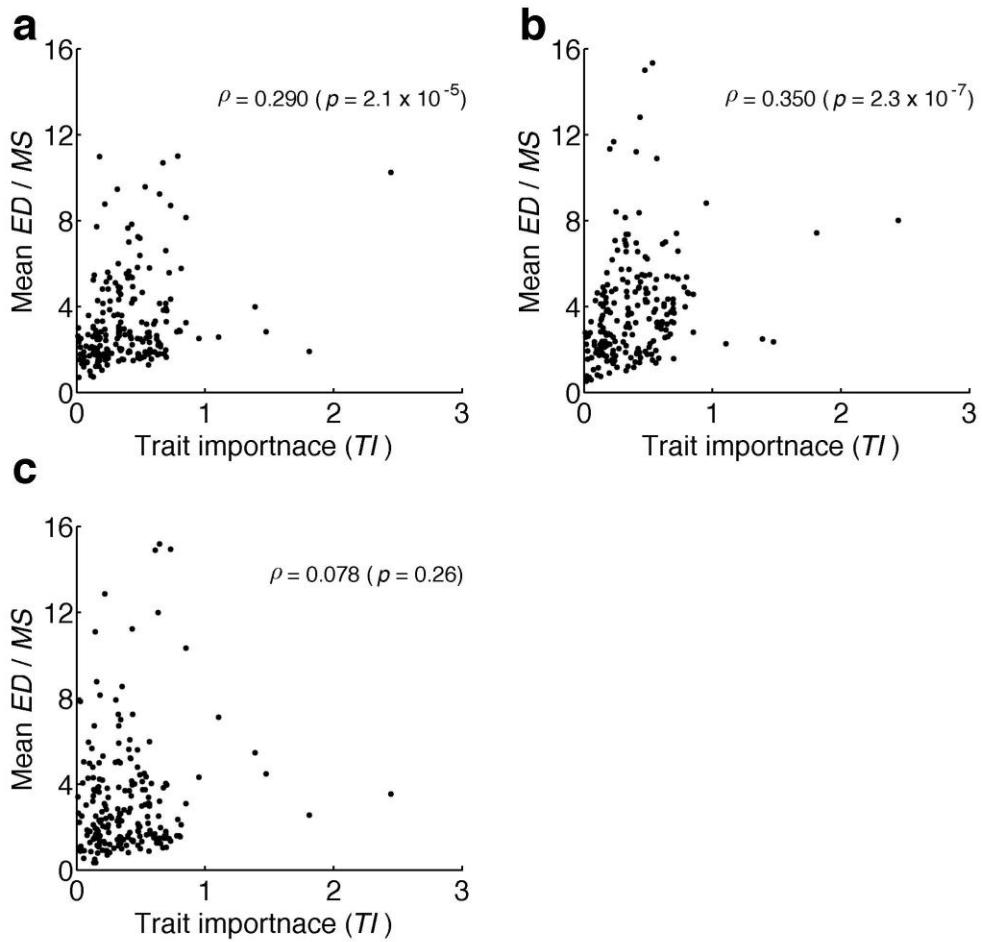


Figure 3