Massively parallel interrogation of the fitness of natural variants in ancient signaling pathways reveals pervasive local adaptation

José Aguilar-Rodríguez[1,2]*, Jean C. C. Vila[1], Shi-An A. Chen[1,3], Manuel Razo-Mejia[1,3], Olivia M. Ghosh[1,4], Daniel F. Jarosz[2,5], Hunter B. Fraser[1], Dmitri A. Petrov[1,6]*

[1]Department of Biology, Stanford University, Stanford, CA, USA

[2]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA

[3]Present address: Altos Labs, Bay Area Institute of Science, Redwood City, CA, USA

[4]Department of Physics, Stanford University, Stanford, CA,

[5]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA,

[6]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA

*To whom correspondence should be addressed: J.A.-R. (jose.aguilar@stanford.edu), D.A.P. (dpetrov@stanford.edu)

Abstract

The nature of standing genetic variation remains a central debate in population genetics, with differing perspectives on whether common variants are almost always neutral as suggested by neutral and nearly neutral theories or whether they can commonly have large functional and fitness effects as proposed by the balance theory. We address this question by mapping the fitness effects of over 9,000 natural variants in the Ras/PKA and TOR/Sch9 pathways—key regulators of cell proliferation in eukaryotes—across four conditions in *Saccharomyces cerevisiae*. While most variants are neutral in our assay, ~3,500 exhibited significant fitness effects. These non-neutral variants tend to be missense and to affect conserved, more densely packed, and less solvent-exposed protein regions. While some of these non-neutral variants are younger and rarer, and more often found in heterozygous states—consistent with purifying selection—a substantial fraction is present at high frequencies in the population, which is expected under balancing selection. Indeed, we find that variants with a positive fitness effect in our laboratory measurement show strong signs of local adaptation as they tend to be found specifically in domesticated strains isolated from human-made environments. Our findings support the view that while many common variants might be effectively neutral, a significant proportion have locally adaptive functional consequences and are driven into a subset of the population by local positive selection. This study highlights the potential to combine high-throughput precision genome editing with fitness measurements to explore natural genetic variation on a pathway-wide scale, thereby bridging the gap between population genetics and functional genomics to understand the nature of evolutionary forces in the wild.

Introduction

One of the longstanding debates in population genetics centers on the nature of standing genetic variation (*1*–*3*). The classical theory (*4*), with nearly neutral theory being its latest incarnation (*5*–*9*), posits that standing genetic variation is generated through the balance between neutral and deleterious mutation and primarily shaped by purifying selection. This model predicts that while variants present at low population frequencies can be both neutral, nearly neutral, or even strongly deleterious, variants present at high frequencies are very unlikely to be deleterious and therefore must be either neutral or nearly neutral. This further implies that common variants are unlikely to have a substantial functional or fitness impacts. This perspective commonly serves as the null model for studies of genetic variation in the wild (*10*).

In contrast, balance theory argues that while purifying selection does keep deleterious alleles at lower frequencies, a meaningful proportion of common genetic variation is functionally important, beneficial in at least a subset of environments, and maintained at high frequencies through the action of balancing selection (*11*). Balancing selection can operate through many mechanisms, including overdominance (heterozygote advantage) (*12*, *13*), frequency-dependent selection (*14*), and selection that varies across space (local adaptation) (*15*) or time (fluctuating selection) (*16*).

Despite much effort and voluminous data, the current population genetic evidence remains inconclusive about which theory best describes population genetic variation (*2*, *3*, *17*). There is clear evidence for most genes that mutations that are more likely to be functional *a priori*, for instance, nonsynonymous variants—those that alter the amino acid sequence of proteins—tend to be present at lower allele frequencies than synonymous variants (*18*–*24*). This implies that they are likely to be disproportionately deleterious. While this pattern is consistent with the classical nearly neutral theory, in the absence of experimental functional data, it is hard

75 to know whether all observed common nonsynonymous or other functional variants are neutral
76 or whether they do contain a substantial fraction of balanced, functionally, and selectively
77 consequential variants.

78 In principle it should be possible to resolve the debate about the functional
79 consequences of the common variation by making direct measurements of the functional
80 significance of individual common variants. However, quantitative genetics approaches, such as
81 quantitative trait loci (QTL) mapping or genome-wide association studies (GWAS), almost
82 universally lack the spatial resolution necessary to determine the functional or fitness
83 consequences of single-nucleotide variants, which are precisely the type of genetic
84 polymorphism most commonly found in natural populations (*20*).

85 Recent advances in precision genome editing allow us to address this question
86 experimentally. While older CRISPR editing methods suffered from low efficiencies (5-20%),
87 limiting the feasibility of massively parallel editing, new techniques in the budding yeast
88 *Saccharomyces cerevisiae* have substantially improved precision genome editing efficiencies
89 (80-100%) (*25*–*30*). This breakthrough now enables the simultaneous introduction of tens of
90 thousands of variants in parallel and the accurate measurement of their fitness *en masse* using
91 genomic barcodes.

92 In this study we have employed one such method, CRISPEY-BAR+, to map the fitness
93 effects of more than 9,000 natural genetic variants with high precision. We chose variants
94 segregating within the core Ras/PKA and TOR/Sch9 pathways, which control cell proliferation in
95 response to environmental stimuli across all eukaryotes (*31*–*33*). These pathways are crucial for
96 the virulence of several pathogenic yeasts (*34*, *35*), and the pathogenesis of diverse human
97 cancers (*36*), and are often mutated to adaptive states in experimental evolution (*37*–*42*). Both
98 standing genetic variation and *de novo* mutations within them contribute to the adaptation to
99 novel nutrient sources in yeast (*37*–*47*), indicating that they may act as a “hot-spot” of

evolutionary novelty and possibly functional genetic variation segregating in this species. The variants we investigate encompass the majority of the known coding and non-coding genetic variation in these pathways across this species (*20*). We therefore built a comprehensive "fitness landscape" (*48*)—a map linking genetic variation to fitness variation—for the standing genetic variants in these pathways. The environments we chose for the interrogation of fitness effects are known to yield large fitness variation for mutations in these pathways (*42*, *46*).

In this study, we find that large-effect functional variants do tend to be disproportionately deleterious, although this relationship is weak. What is surprising and inconsistent with nearly neutral theory is that many very large-effect polymorphisms are present at high frequencies in the population. Moreover, we find that variants that have positive fitness effects in our laboratory measurements are enriched in domesticated and industrial strains, suggesting that they have been driven to high frequency by local positive selection during adaptation to fermented food and other human-made environments (*49*). These results provide strong support in favor of the balance theory of natural variation. This study not only highlights the potential of high-throughput precision genome editing to resolve foundational questions in population genetics, but also to expand the exploration of empirical fitness landscapes beyond single-gene approaches toward entire biological pathways (*50*, *29*). By focusing on *natural* rather than *random* mutations, and by integrating ecological, phylogenetic, and allele-frequency information, our approach makes it possible to connect high-throughput laboratory fitness measurements to patterns of genetic variation in nature in a way that has not been possible before.

125 **Results**

126 ***High-throughput precision genome editing enables measurement of fitness effects for***
127 ***thousands of natural variants in the laboratory***

128 To quantify the fitness effects of individual genetic variants in the Ras/PKA and TOR/Sch9
129 pathways we employed CRISPEY-BAR+, a modified version of a previously published high-
130 throughput CRISPR-based precision genome editing technology. This previous method, called
131 CRISPEY (*26*), provides a means to introduce many predefined mutations into a yeast
132 population, one per cell, through a plasmid library of guide RNAs and corresponding donor
133 templates for each. In this system, the donor DNA is generated *in vivo* by a bacterial retron
134 element, which reverse-transcribes the donor RNA into single-stranded DNA that serves as a
135 repair template for Cas9-induced double-stranded breaks. As a result of the unique
136 biochemistry of retrons, the donor DNA is covalently linked to the RNA molecule that contains
137 the guide RNA, ensuring its co-localization with the Cas9/gRNA complex. Critically, because the
138 method has such high editing efficiency, it is possible to infer the frequency of the edit within a
139 population by measuring the abundance of the corresponding guide/donor, thereby enabling
140 pooled measurements of fitness via high-throughput sequencing of the plasmids. An
141 improvement of the original method, CRISPEY-BAR, is based on a new vector that incorporates
142 two consecutive retron-guide-donor cassettes, allowing the simultaneous generation of two
143 guide/donor pairs for making two precise edits in the same cell (*29*, *30*). One guide/donor pair
144 integrates a genomic barcode in a neutral locus, and the other introduces a precise edit of
145 interest elsewhere in the genome. With our new method CRISPEY-BAR+ we also generate one
146 precise edit and a unique barcode per cell and, in a pooled assay, measure the change in
147 abundance of the edited strain through deep sequencing of the chromosomal region where
148 barcodes are integrated (Fig. 1A). However, in contrast to CRISPEY-BAR (*29*, *30*), every
149 guide/donor pair, and thus the variants that they introduce, is associated with hundreds of

150 unique barcodes (Fig. S1A), providing robust internal controls to ensure changes in fitness are
151 linked to genome editing and improving the precision of our subsequent fitness measurement
152 (Fig. S1B). The association between barcodes and edits is established by sequencing the
153 plasmid library (final plasmid library in Figure 1A; Methods).

154 After curating the relevant literature, we selected 18 genes from the Ras/PKA pathway
155 and 6 from the TOR/Sch9 pathway (Supplementary Table S1), focusing on those most central
156 to these signaling pathways. Using genome-wide polymorphisms identified from whole-genome
157 sequences of 1,011 genetically and ecologically diverse yeast isolates (spanning wild and
158 domesticated strains) (*20*), we identified 14,213 variants segregating in *S. cerevisiae* within
159 these pathways, including both single-nucleotide polymorphisms (SNPs) as well as indels. Most
160 (9,953) were in one of the 24 protein-coding regions, while the remaining 4,260 were in adjacent
161 noncoding regions.

162 We were able to target 9,589 of those 14,213 variants with CRISPEY-BAR+, based on
163 their proximity to a PAM site required by Cas9. Across edits, we observed an average editing
164 frequency of ~85% determined from randomly picked barcoded strains (Methods), which is in
165 agreement with previous studies using the CRISPEY editing technology (*28*, *29*). The 9,589
166 variants are broadly distributed across the 1,011 strains (Fig.1B), and span a wide range of
167 functional types: 37.9% (3,629) are nonsynonymous, 36.2% (3,473) are synonymous, 24.5%
168 (2,349) are noncoding, and the remaining 1.4% (138) are either frameshifts, insertions/deletions
169 in coding regions, or impacted start/stop codons.

170 To investigate fitness in an environment linked to the function of the Ras/PKA and
171 TOR/Sch9 pathways, we competed the edited pool in a minimal medium with limiting glucose
172 (1.5%, M3) over ~25 generations in serial batch culture (*51*). We conducted growth competitions
173 under four experimental conditions, each with three technical replicates. The experimental
174 conditions varied in the duration the cells spent in the flask before transfer: 24 hours (1D), 48

175 hours (2D), 72 hours (3D), and 120 hours (5D) (Fig. 1C). As the growth cycle lengthens, cells
176 transition from active fermentation (1D) to fermentation and respiration (2D) to prolonged
177 stationary phase at 3D and 5D, once the limiting nutrients are depleted. Previous work with
178 adaptive mutants in the Ras/PKA and TOR/Sch9 pathways, conducted in laboratory evolution
179 experiments, demonstrated that these growth conditions should enable us to reveal the fitness
180 effects of functional variation in these pathways (*42*, *46*).

181 During the growth competitions, we sampled cultures every ~4.3 generations. We
182 employed high-throughput sequencing of the barcode region in these samples to monitor the
183 relative abundance of variants over time (Fig. 1D). We then used these frequency trajectories to
184 estimate fitness effects of each variant. To control for the distribution of neutral fitness
185 measurements that arise in wild-type cells from experimental noise and genetic drift, we
186 measured variant fitness effects from each competition relative to non-editing control plasmids
187 that do not alter the genome apart from the barcode insertion (Fig. 1D). We used a Bayesian
188 approach to quantify the selection coefficients for each variant (*52*). This approach is based on
189 a hierarchical model that employs all experimental replicates simultaneously for the inference of
190 fitness (Methods). We estimated selection coefficients for 9,447 variants out of the 9,589 that
191 we could target with CRISPEY-BAR+. The remaining 142 variants lacked sufficient barcodes in
192 the competition data to assess their fitness. We found that the fitness effects measured by
193 CRISPEY-BAR+ were highly reproducible across replicates, with an average Pearson's r of
194 0.93 across conditions and pairwise replicate comparisons ($P < 2.2 \times 10^{-16}$; Fig. 1E and Fig.
195 S2). To validate that observed fitness effects reflect intended edits, we analyzed control
196 nonsense edits in cysteine or tyrosine codons of genes known to impair fitness when deleted
197 (*26*). These nonsense edits showed significantly stronger fitness defects than synonymous edits
198 at the same positions, supporting the accuracy and specificity of genome editing in our assay
199 (Wilcoxon's rank test, $P = 6.52 \times 10^{-36}$; Fig. S3A).

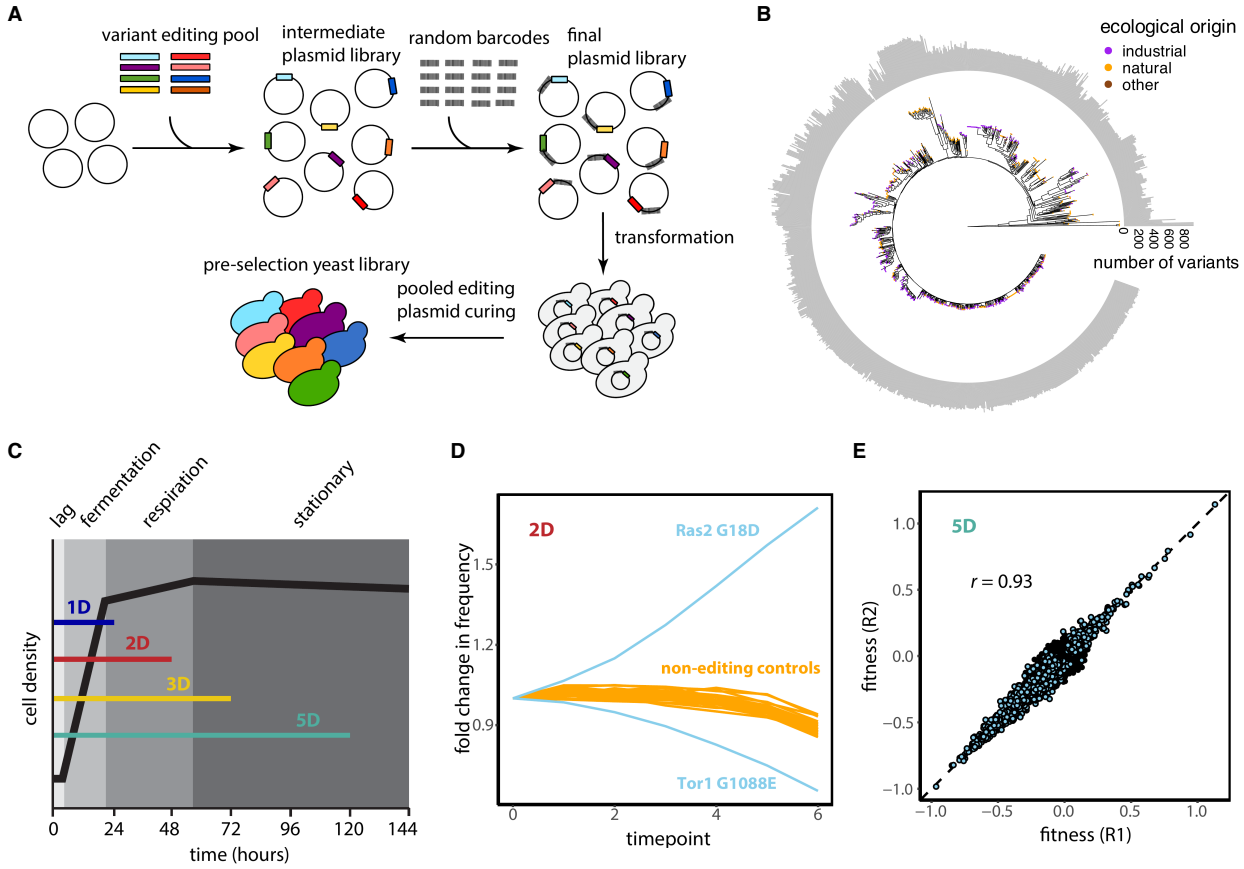


Figure 1. **CRISPEY-BAR+ fitness screen of natural genetic variants in nutrient-sensing pathways.** (A) Schematic of the CRISPEY-BAR+ approach: from cloning to pre-selection yeast population. The association between random barcodes and editing oligos is obtained by sequencing the final plasmid library. (B) Maximum likelihood tree of the 1,011 yeast strains. Individual strains are labeled according to their ecological origin, based on the provenances reported by Peter *et al.* (*20*), as industrial, natural, or other (see Methods). We report the number of variants in the Ras/PKA and TOR/Sch9 pathways for each strain. (C) Schematic of the four experimental conditions with varying time before transfer (1D: 24 h, 2D: 48 h, 3D: 72 h, 5D: 120 h) to change the amount of time spent in different phases of the growth cycle (lag, fermentation, respiration, and stationary). (D) Example of competition data over time from the 2D condition. Each blue line represents the average across three experimental replicates of the normalized counts for two missense edits: a beneficial (Ras2 G18D) and a negative (Tor1 G1088E). The beneficial variant is adjacent to the most common G12X oncogenic mutation in humans (*RAS2* is an ortholog of *KRAS, NRAS, and NRAS*) (*53*). Counts in later time points are normalized to the counts in the first time

point. All the barcodes mapping to the same edit are counted. Orange lines represent the normalized counts for barcodes associated with non-editing oligos, representing a neutral fitness baseline. These orange lines go down in frequency because as the mean fitness of the population increases as a consequence of the increase in frequency of beneficial variants. (F) Estimates of relative fitness per cycle are highly reproducible across technical replicates. The scatter plot shows fitness values across replicate 1 (R1) and replicate 2 (R2) for each variant in the 5D condition. Non-neutral variants—those with significant fitness effects—are highlighted in blue, while neutral variants are shown in black. The dashed diagonal line indicates the identity line, and the Pearson correlation coefficient (*r*) quantifies replicate agreement. See Figure S2 for all replicate comparisons across all conditions.

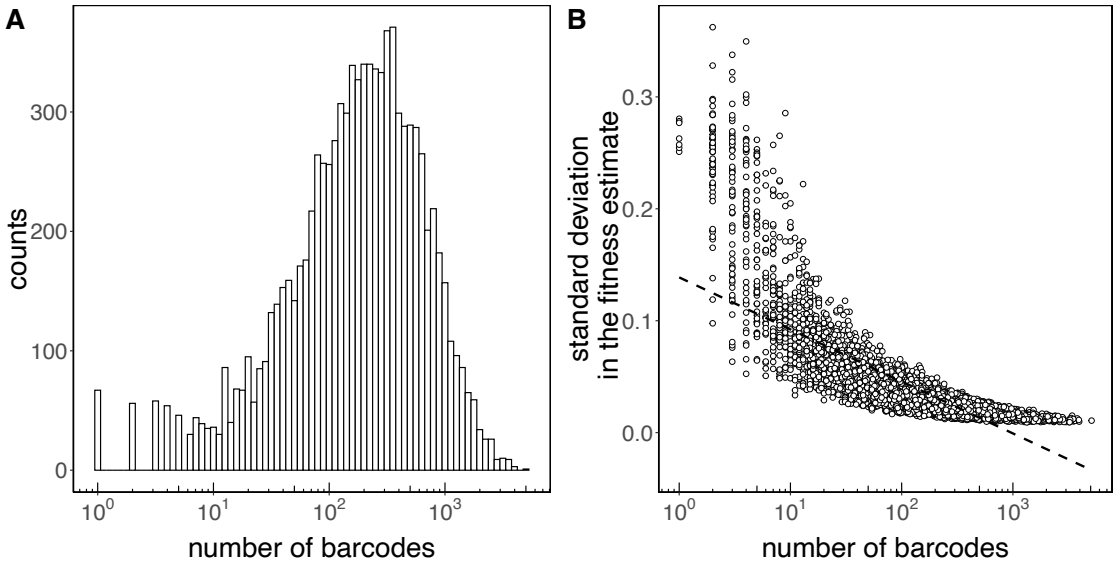


Figure S1. Barcode coverage and its relationship to measurement precision. (A) Histogram showing the number of barcodes per edit for all edits at timepoint = 0 of the pooled growth competitions (Methods). (B) For the same edits, standard deviation of the fitness estimate in the 2D condition as a function of the number of barcodes. A dashed line shows the best least-squares fit to the data provided as a visual guide. The x-axis is displayed on a $\log_{10}$ scale in both panels. See Methods and the first section of Results for details on pooled growth competitions, experimental conditions, and fitness estimation.

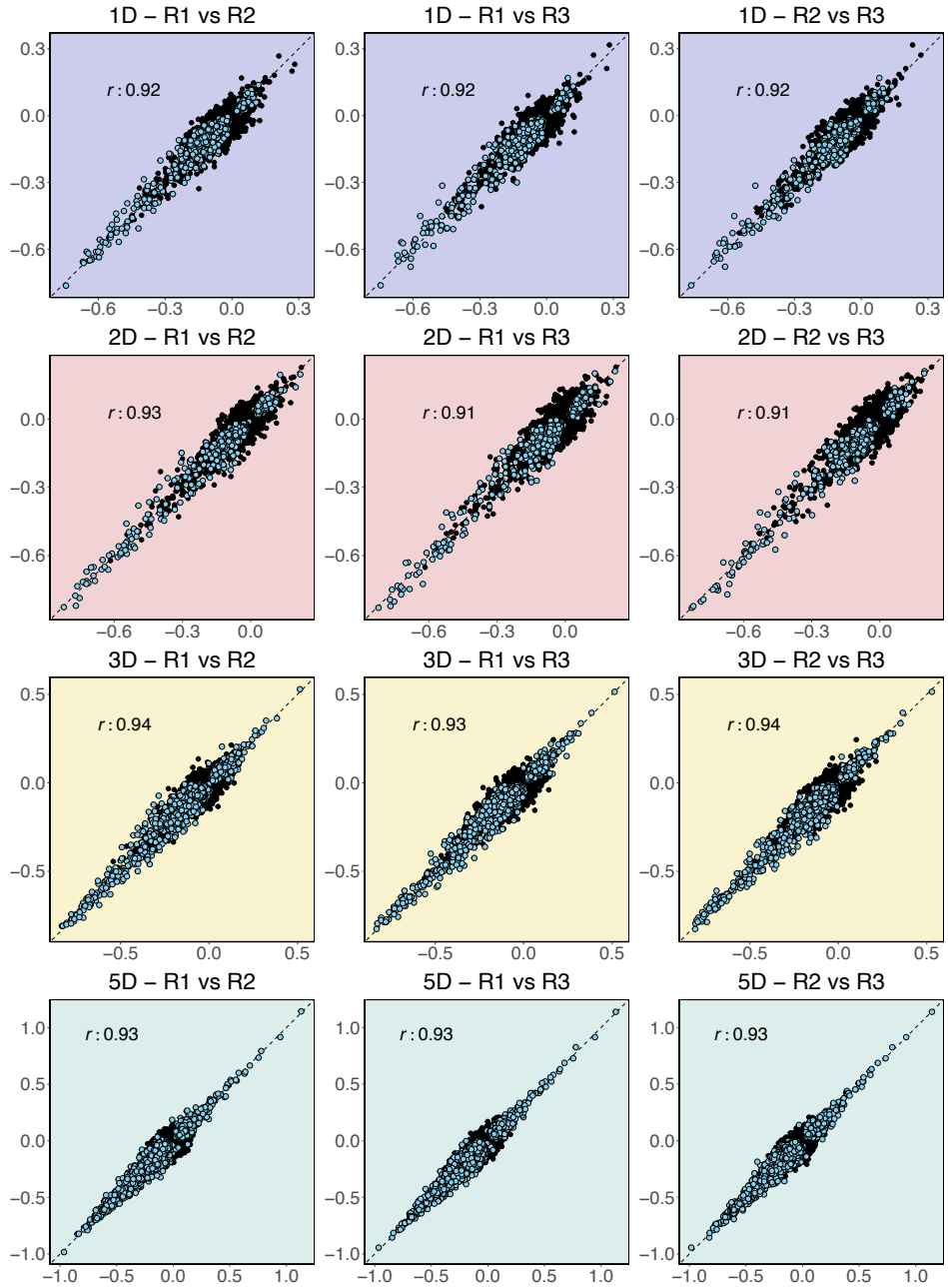


Figure S2. Fitness estimates are highly reproducible across technical replicates in the four conditions. Scatter plots show pairwise comparisons of fitness values for each variant across replicates (R1, R2, and R3) in each condition (1D, 2D, 3D, 5D). Neutral variants are shown in black, while non-neutral variants—those with significant fitness effects—are highlighted in blue. Dashed diagonal lines indicate the identity line, and Pearson's correlation coefficients (*r*) quantify the agreement between replicates.

250 ***CRISPEY-BAR+ reveals widespread fitness effects of natural variants in Ras/PKA and***
251 ***TOR/Sch9 signaling***

252 To identify variants with fitness effects in our laboratory conditions, we classified a variant as
253 significantly non-neutral if the 95% posterior credible interval for the estimated selection
254 coefficient excluded zero (*52*). Using this criterion, we identified 3,673 (39%) variants that had
255 significant fitness effects in at least one condition (Fig. 2A,B). Of these, 659 variants were
256 beneficial in at least one environment (Fig. 2C,D), while the remainder were exclusively
257 deleterious when non-neutral. Notably, 78 beneficial variants also exhibited significant adverse
258 fitness effects in at least one other condition (Fig. 2C), revealing fitness trade-offs across
259 different growth phases. While we refer to these variants as beneficial or deleterious here, we
260 note that because yeast populations experience diverse and fluctuating environments in the
261 wild, a variant beneficial under laboratory-like conditions may be mostly neutral or even
262 deleterious in nature, depending on how frequently a lineage experiences a dissimilar type of
263 environment. The smallest absolute fitness effect that we detected with this approach was ~2%
264 per transfer, providing a lower bound on the effect sizes that we can reliably resolve. The
265 number of variants with significant fitness effects increased with longer growth cycles, peaking
266 in the 5D condition (Fig. S3B).

267 Non-neutral variants are widespread across both the Ras/PKA and TOR/Sch9 pathways,
268 with no entire gene devoid of them (Fig. 2E and Fig. S5). To assess whether specific genes
269 carry more non-neutral variants than expected given the number of assayed variants on them,
270 we performed binomial tests, accounting for the overall background rates of fitness-altering
271 variation within each pathway (Benjamin-Hochberg false discovery rate, FDR < 0.05). We
272 identified three genes with a significant excess of non-neutral variants (Fig. 2F): *CYR1,* which
273 encodes the cAMP-generating adenylate cyclase; *TPK3*, a subunit of the cAMP-dependent
274 protein kinase A (PKA); and *LST8*, a regulatory subunit shared by both the Tor1-containing

275 TORC1 and the Tor2-containing TORC2 complexes. When separating non-neutral variants
276 further by beneficial or deleterious, we found that beneficial variants were particularly enriched
277 in the terminal effector kinases of each pathway (Fig. S4A): kinase Sch9 (*SCH9*) in the
278 TOR/Sch9 pathway and PKA (*TPK3*) in the Ras/PKA pathway. These two genes (*SCH9* and
279 *TPK3*) alone account for approximately 12% of all beneficial variants detected (Fig. 2E). This
280 enrichment may reflect the fact that these kinases are positioned at the downstream ends of
281 their respective signaling cascades, where they directly control growth-related transcriptional
282 and metabolic programs. As terminal effectors, mutations in these genes may offer a fine-tuned
283 or conditionally advantageous means of modulating pathway output, potentially making them
284 frequent targets of adaptation.
285
286
287
288
289
290
291
292
293
294
295

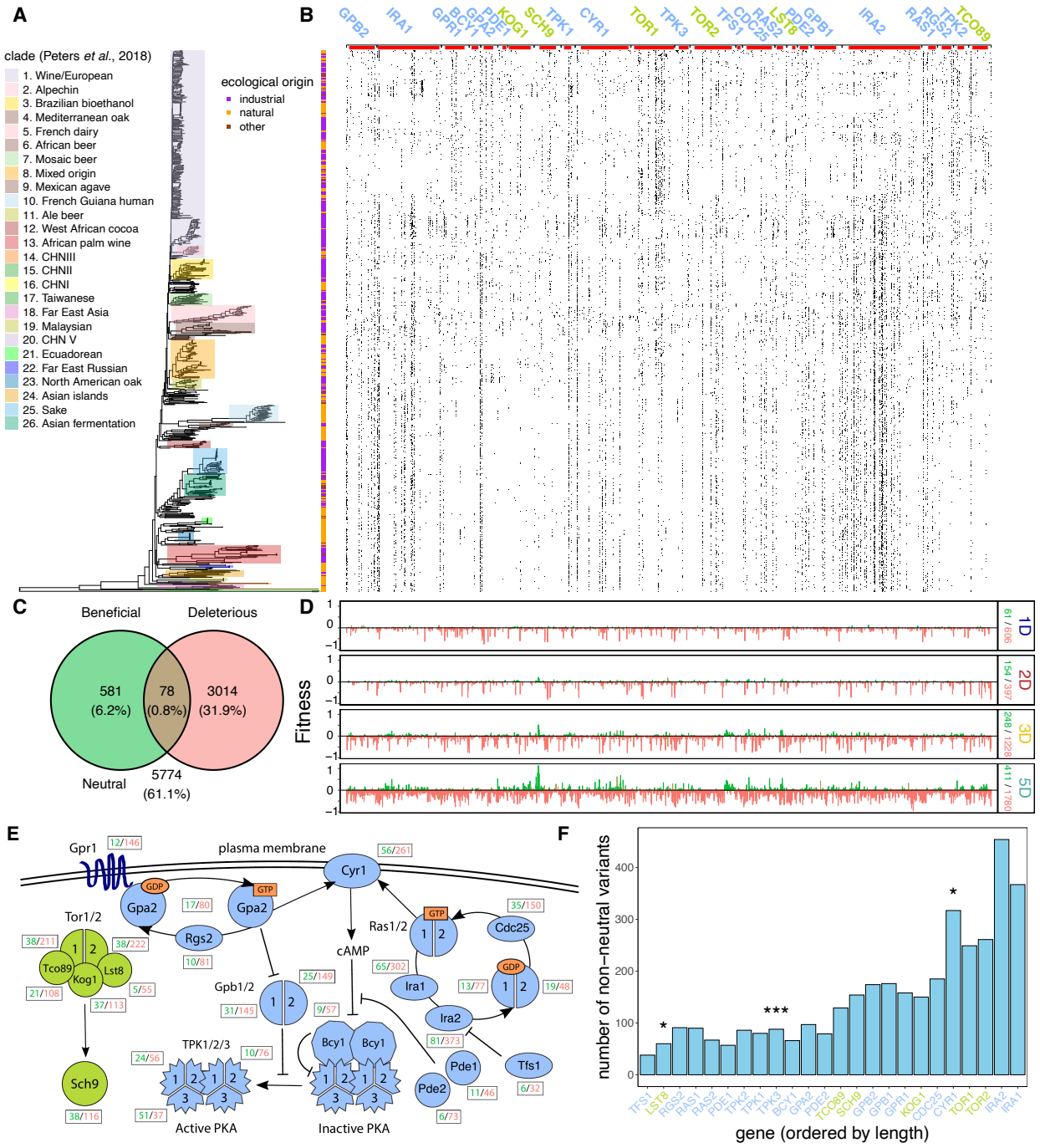


Figure 2. Fitness effects of segregating genetic variants in nutrient-sensing pathways revealed by CRISPEY-BAR+. (A) Maximum likelihood tree depicting phylogenetic relationship between 1,011 yeast strains (Methods). Monophyletic clades are labeled according to Peters *et al.* (*20*) and individual strains are classified according to their ecological provenance as either industrial or natural (Methods). (B) Matrix

301 showing the presence and absence of non-neutral variants in each of the 1,011 yeast strains in the tree of
302 panel A. Each row corresponds to a strain, and each column to a variant, ordered by genomic position.
303 Gene annotations are indicated above the matrix, and the coding region of each gene is highlighted as a
304 red bar. Gene names are colored by pathway: blue for Ras/PKA and green for Tor/Sch9 (following the
305 color scheme in panel E). (C) Venn diagram showing the number of non-neutral variants that are
306 significantly beneficial in at least one condition, significantly deleterious in at least one condition, or
307 significantly beneficial and deleterious in different conditions (intersection). (D) Fitness of all non-neutral
308 variants in each of the four conditions tested. Variants are colored according to panel C. The number of
309 non-neutral variants that are beneficial (green) or deleterious (red) in each condition is listed. (E)
310 Schematic of the signaling pathways Ras/PKA (blue) and TOR/Sch9 (green) in yeast. Positive signaling
311 interactions are depicted with standard arrowheads and negative (inhibitory) interactions with flat-head (T-
312 bar) arrows. The number of non-neutral variants that are beneficial (green) or deleterious (red) in each
313 gene is listed. (F) Bar plot showing the number of non-neutral variants per gene, ordered by increasing
314 gene length. Gene names are colored by pathway: blue for Ras/PKA and green for Tor/Sch9 (following
315 the color scheme in panel E). Asterisks indicate genes with a statistically significant excess of variants
316 based on a binomial test corrected for multiple testing (*FDR < 0.05; **FDR < 0.01; ***FDR < 0.001). See
317 Figure S4 for the corresponding distribution of beneficial and deleterious variants among these pathway
318 genes.

319

320

321

322

323

324

325

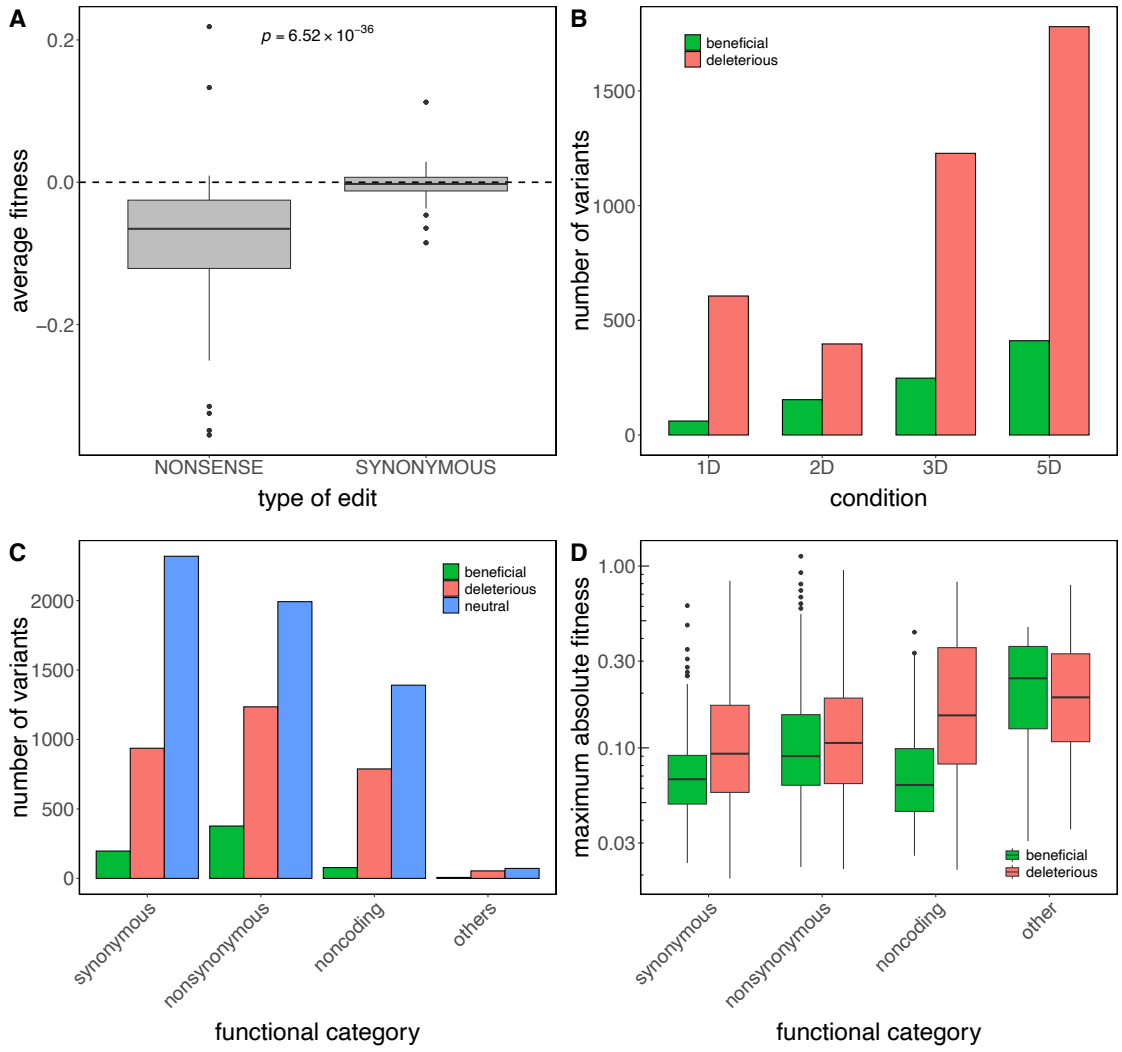


Figure S3. (A) Box plot of average fitness effects across the four tested conditions for edits modifying cysteine or tyrosine codons into stop codons, and edits introducing synonymous substitutions in the same codons. (B) Bar plot of the number of beneficial and deleterious variants by experimental condition (1-Day condition, 2-Day condition, 3-Day condition, and 5-Day condition). (C) Bar plot of the number of neutral, deleterious, and beneficial variants by variant effect. (D) Box plot of the maximum absolute value of fitness effects across the four conditions tested by variant effect for beneficial and deleterious variants.

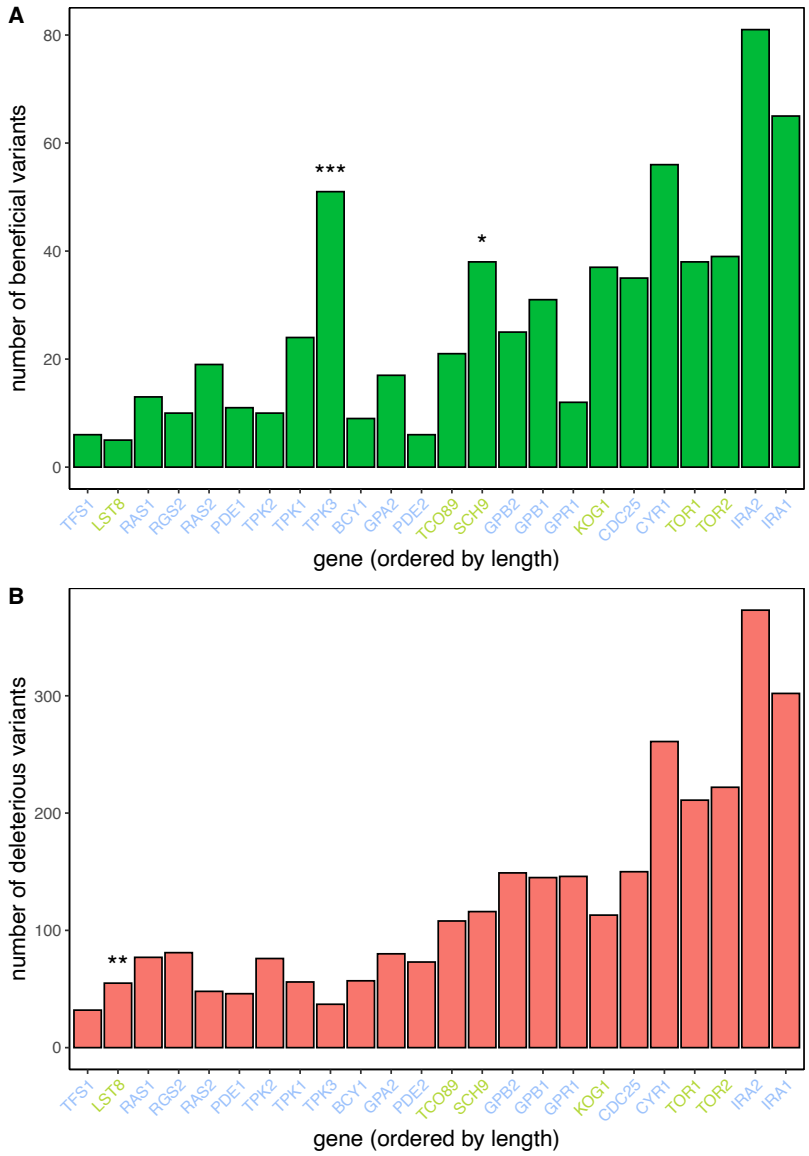


Figure S4. Variant burden across signaling genes in the Ras/PKA and Tor/Sch9 pathways. Bar plots show the number of beneficial (panel A) and deleterious (panel B) variants per gene, ordered by increasing gene length. Bars are colored by variant type: green for beneficial, and red for deleterious variants. Gene names are colored by pathway: blue for Ras/PKA and green for Tor/Sch9 (following the color scheme in Fig. 2E). Asterisks indicate genes with a statistically significant excess of variants based on a binomial test with multiple-testing correction applied at the pathway level (*FDR < 0.05; **FDR < 0.01; ***FDR < 0.001).

342 ***Non-neutral variants tend to be nonsynonymous and are predicted to be functional***

343 We next sought to investigate the predicted molecular impacts of the variants that show
344 significant fitness effects in our CRISPEY-BAR+ measurements, leveraging the single-
345 nucleotide resolution of our approach. Non-neutral variants are widely distributed across all
346 functional categories of variants (Fig. S3C). In line with prior expectation of functionality, large
347 fractions (~60%; 25/42) of variants that abolish start/stop codons or create new stop codons,
348 and ~44% (1,612/3,629) of all non-synonymous variants are significantly non-neutral, in both
349 cases, this is a significantly higher proportion than either synonymous (~33%; 1,134/3,473) or
350 noncoding (~37%; 866/2,349) variants, which are generally less likely to impact function
351 (Fisher's exact test, $P << 0.001$ in all cases; Fig. S3C). Moreover, the overrepresentation of
352 nonsynonymous variants amongst non-neutral variants increases with the number of conditions
353 in which a variant is non-neutral (Fig. 3A, Fig. S5A, and Fig. S6A), and non-synonymous
354 variants have on average substantially larger fitness effects than synonymous variants
355 (Wilcoxon's test, $P = 1.7 \times 10^{-7}$; Fig. S3D). Consistent with a previous study (*54*), non-neutral
356 synonymous variants tend to decrease codon optimality specifically near the start of genes
357 relative to neutral variants (Fig. S7). This may reflect changes to the *ribosome ramp*—a region
358 of suboptimal codons near the 5' end of coding sequences thought to reduce ribosome
359 crowding and optimize translation efficiency (*54*).

360 We then proceeded to evaluate how specifically nonsynonymous non-neutral variants
361 affect different protein properties. To do so, we related the number of conditions in which a
362 nonsynonymous variant is non-neutral to functional predictions based on evolutionary
363 conservation or protein structure and stability (Fig. 3B-E, Fig. S5B-E, and Fig. S6B-E). Variants
364 that are non-neutral in multiple environments are expected to have larger, more robust
365 molecular consequences, rather than condition-specific or weak effects.

366 We first found that if a missense variant occurs within a conserved protein region it is
367 more likely to have an impact on fitness. For this analysis, we used Sorting Intolerant from
368 Tolerant (SIFT) scores, which predict whether an amino acid substitution is likely to affect
369 protein function based on sequence conservation and the physical properties of the amino acids
370 involved (*55*, *56*). We find that non-neutral nonsynonymous variants tend to have lower SIFT
371 scores than neutral nonsynonymous variants suggesting they are more likely to affect protein
372 function (Wilcoxon's test, $P = 1.1 \times 10^{-7}$; Fig. 3B, Fig. S5B, and Fig. S6B). Variants with a SIFT
373 score below 0.05 are considered to have particularly strong functional impacts (*55*). There is an
374 enrichment of such variants among non-neutral variants in our assays (Fisher's test, $P = 1.6 \times$
375 10^{-8}).
376 We next examined three key protein properties: packing density, solvent accessibility,
377 and protein stability. Packing density reflects how closely a residue is located to others within
378 the protein's tertiary structure, essentially measuring the degree of contact a residue has with
379 other residues. Solvent accessibility represents the surface area of a residue that is exposed to
380 water. Substituting residues in densely packed regions, as well as those with lower solvent
381 exposure, is likely to induce considerable changes to the protein's structure and are therefore
382 more evolutionarily constrained (*57*). Indeed, we find that nonsynonymous variants that are non-
383 neutral in more environments tend to affect residues with higher packing density (Fig. 3C, Fig.
384 S5C, and Fig. S6C) and those that are less solvent-exposed (Fig. 3C, Fig. S5C, and Fig. S6C).
385 As a control, we confirmed that there were no significant differences in packing density or
386 solvent accessibility comparing codons containing non-neutral relative to those containing
387 neutral synonymous variants (Fig. S8). Because nonsynonymous variants in these structurally
388 constrained sites often destabilize the folded protein, we next asked whether non-neutral
389 variants are more likely to impair stability. Using FoldX to estimate changes in folding free

energy (ΔΔG) (*58*), we found an enrichment of strongly destabilizing variants (ΔΔG > 2) among those that are non-neutral in multiple environments (Fig. 3E, Fig. S5E, and Fig. S6E).

In summary, our findings indicate that non-neutral variants are more likely to be missense mutations, affecting conserved, densely packed, and buried residues within a protein, and often resulting in destabilization. These results provide a strong biochemical basis for the distinction between neutral and non-neutral variants, further validating the findings of our high-throughput fitness measurements.

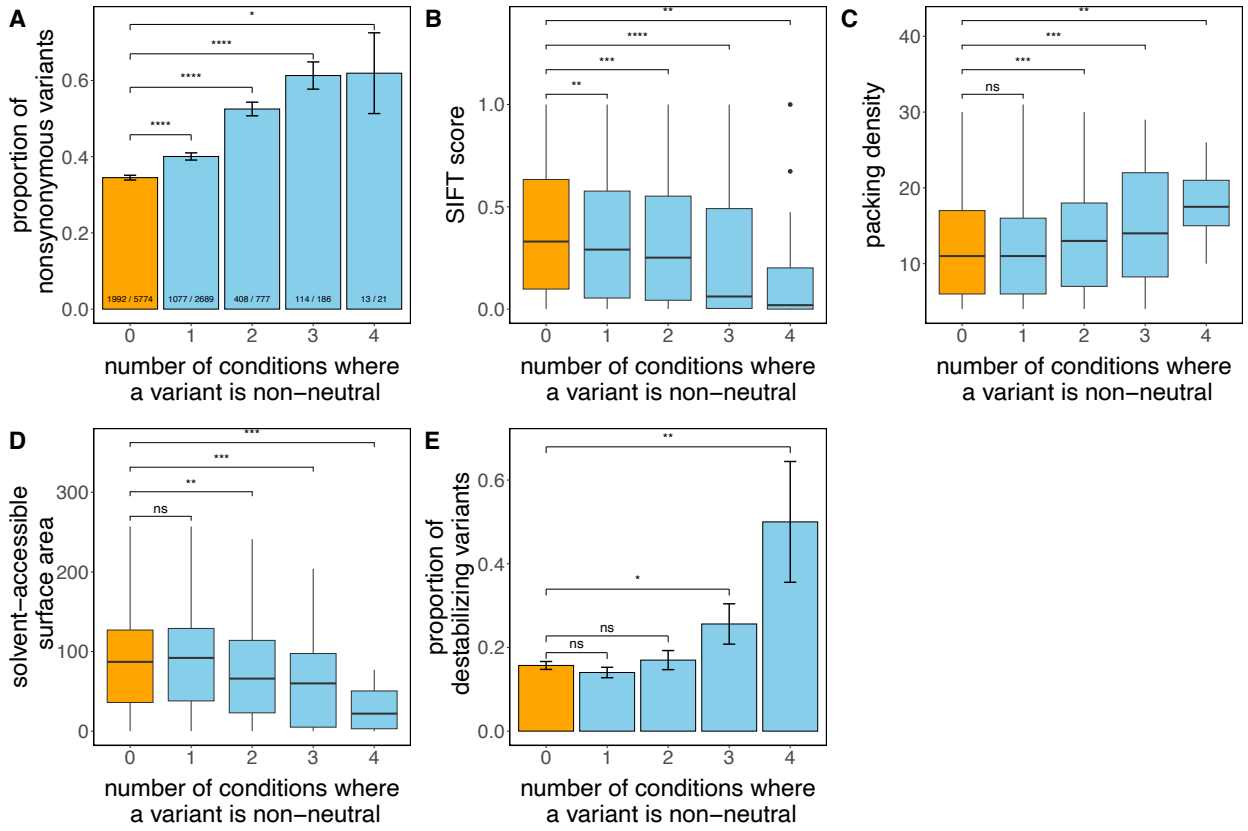


Figure 3. **Non-neutral variants tend to be missense and are predicted to be functional.** (A) Bar plot showing an enrichment of nonsynonymous variants among non-neutral variants compared to neutral variants. In this and all subsequent panels, non-neutral variants are grouped by the number of conditions in which they exhibit non-neutral fitness effects. The ratios indicate the number of nonsynonymous variants among all the variants of each class. (B) Box plot showing that non-neutral variants tend to affect more evolutionarily conserved residues, as indicated by lower SIFT scores. (C) Box plot showing that non-neutral variants tend to affect residues with higher packing density. (D) Box plot showing that non-neutral variants tend to affect residues that are less exposed to the solvent. (E) Proportion of destabilizing variants for non-neutral and neutral variants. Differences between non-neutral and neutral variants were assessed using Fisher's exact test for panels A and E, and the Wilcoxon test for the remaining panels. Significance levels are indicated as follows: $P \leq 0.0001$: ****, $P \leq 0.001$: ***, $P \leq 0.01$: **, $P \leq 0.05$: *, not significant: ns. Box plots show the median an interquartile range; whiskers extend to 1.5 times the interquartile range.

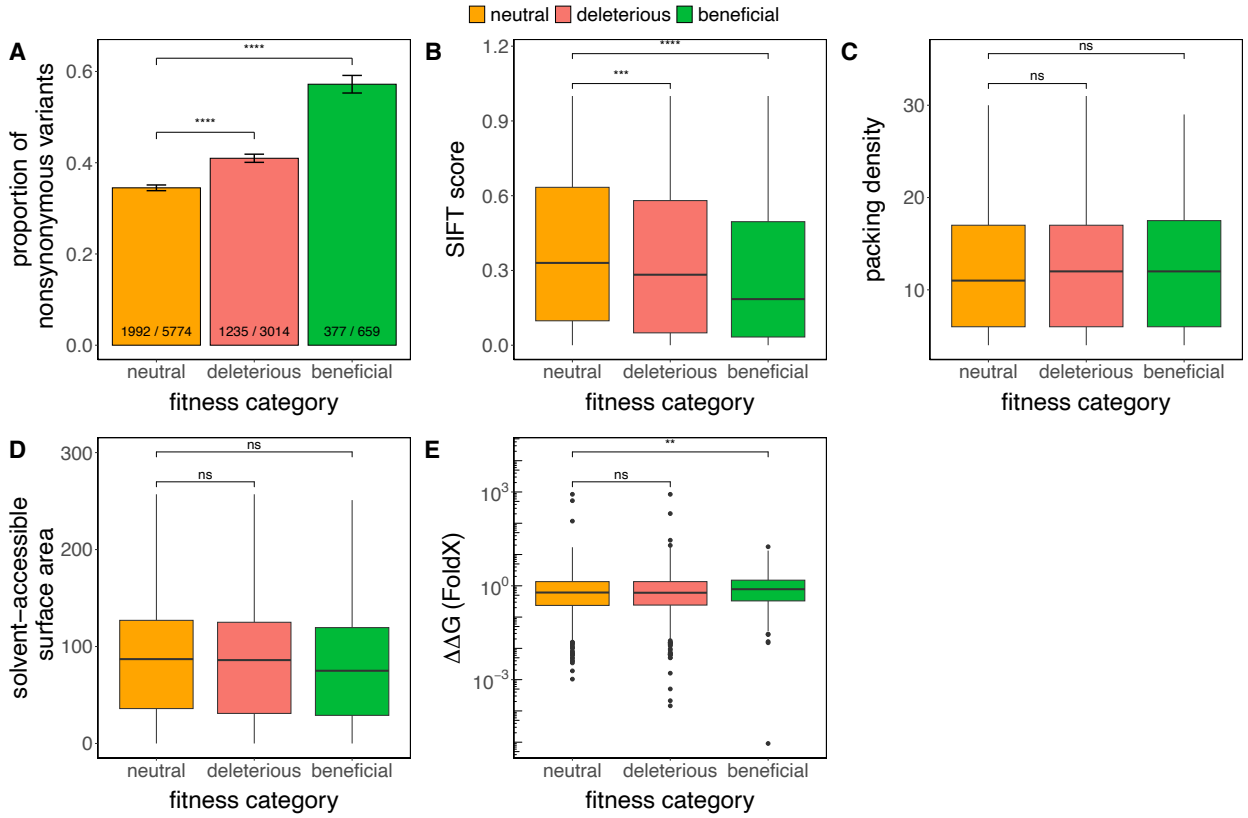


Figure S5. **To accompany Figure 3. Non-neutral variants tend to be missense and are predicted to be functional.** (A) Bar plot showing an enrichment of nonsynonymous variants among non-neutral variants compared to neutral variants. In this and all subsequent panels, non-neutral variants are grouped as beneficial if they have a positive fitness effect in at least one condition, or as deleterious otherwise. The ratios indicate the number of nonsynonymous variants among all the variants of each class. (B) Box plot of SIFT scores for neutral, deleterious, and beneficial variants. (C) Box plot of packing density for neutral, deleterious, and beneficial variants. (D) Box plot of solvent accessibility for neutral, deleterious, and beneficial variants. (E) Box plot of ΔΔG as computed by FoldX for neutral, deleterious, and beneficial variants. Differences between non-neutral and neutral variants were assessed using Fisher's exact test for panel A, and the Wilcoxon test for the remaining panels. Significance levels are indicated as follows: $P \leq 0.0001$: ****, $P \leq 0.001$: ***, $P \leq 0.01$: **, $P \leq 0.05$: *, not significant: ns. Box plots show the median an interquartile range; whiskers extend to 1.5 times the interquartile range.

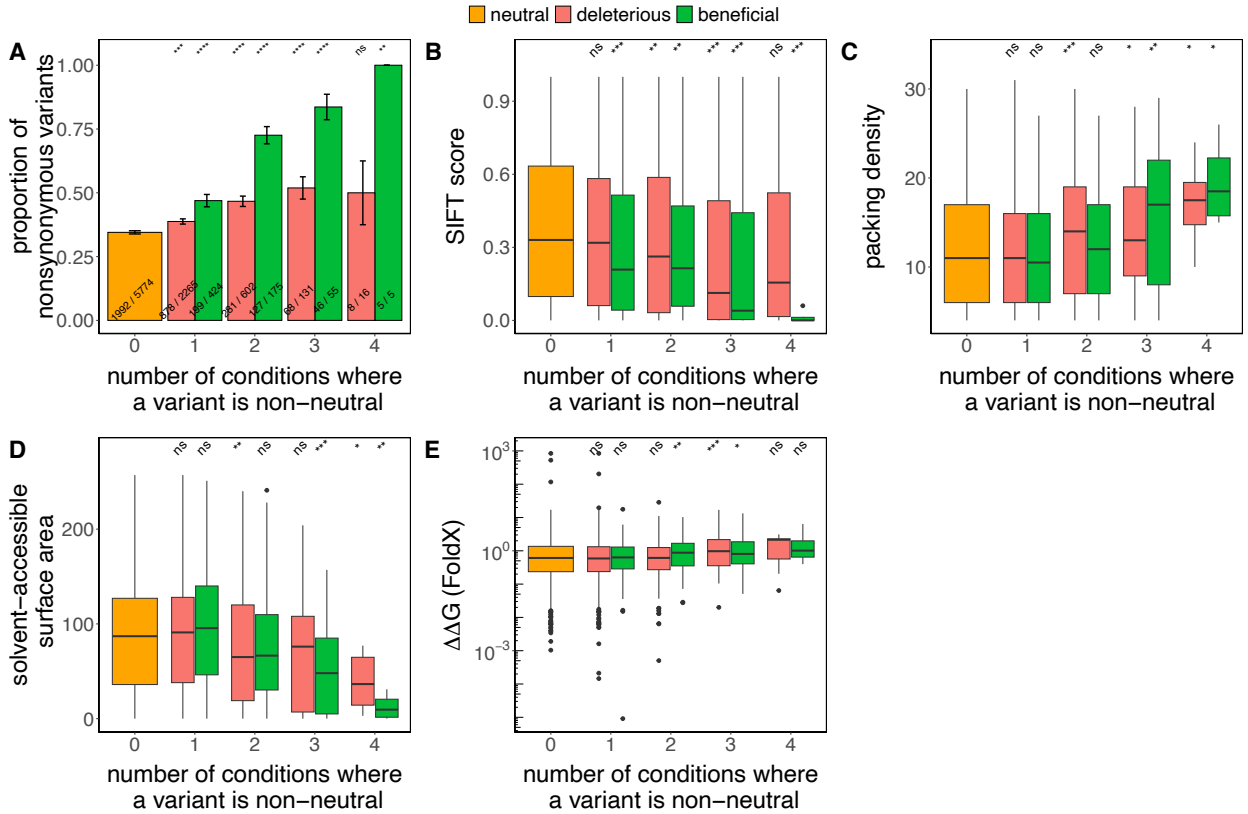


Figure S6. **To accompany Figure 3. Non-neutral variants tend to be missense and are predicted to be functional.** (A) Bar plot showing an enrichment of nonsynonymous variants among non-neutral variants compared to neutral variants. In this and all subsequent panels, non-neutral variants are grouped as beneficial if they have a positive fitness effect in at least one condition, or as deleterious otherwise, and by the number of conditions in which they exhibit non-neutral fitness effects. The ratios indicate the number of nonsynonymous variants among all the variants of each class. (B) Box plot of SIFT scores for neutral, deleterious, and beneficial variants. (C) Box plot of packing density for neutral, deleterious, and beneficial variants. (D) Box plot of solvent accessibility for neutral, deleterious, and beneficial variants. (E) Box plot of ΔΔG as computed by FoldX for neutral, deleterious, and beneficial variants. Differences between non-neutral and neutral variants were assessed using Fisher's exact test for panel A, and the Wilcoxon test for the remaining panels. Significance levels are indicated as follows: $P \leq 0.0001$: ****, $P \leq 0.001$: ***, $P \leq 0.01$: **, $P \leq 0.05$: *, not significant: ns. Box plots show the median an interquartile range; whiskers extend to 1.5 times the interquartile range.

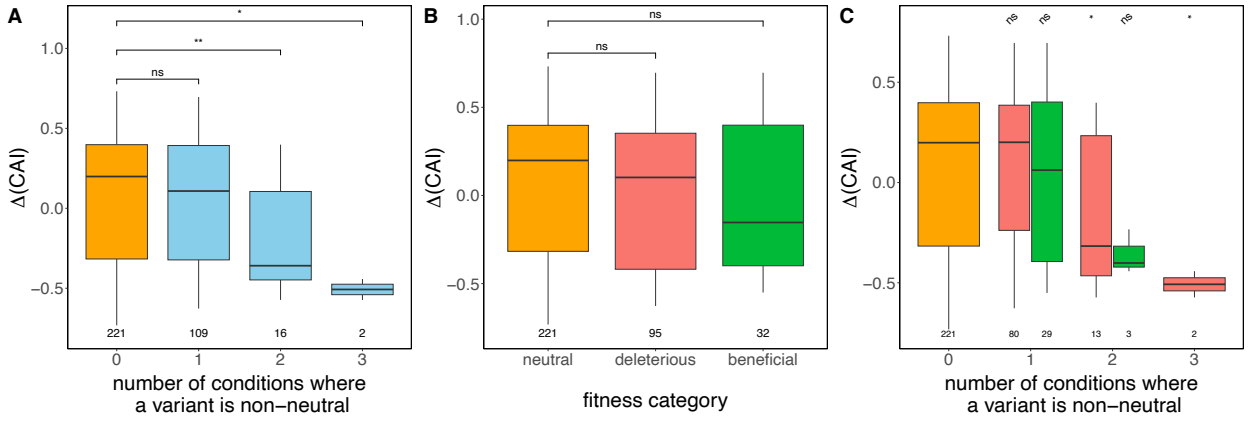


Figure S7. **Synonymous variants with fitness effects tend to decrease the codon adaptation index (CAI) when located near the *N*-terminus.** (A) Box plots showing the change in codon adaptation index (ΔCAI) between reference and variant codons for synonymous variants located within the first 10% of the protein. Variants are grouped by the number of experimental conditions in which they exhibit significant non-neutral fitness effects. (B) The same ΔCAI values for these variants are now grouped by fitness effect category (beneficial, neutral, or deleterious). (C) ΔCAI values for the same synonymous variants, grouped by both fitness effect category and the number of conditions in which the variant is non-neutral. Box plot center lines show medians, boxes indicate interquartile ranges, and whiskers extend 1.5 times the interquartile range. Sample sizes are indicated below each group. Statistical significance was assessed using pairwise Wilcoxon rank-sum tests; P values are shown as follows: $P \leq 0.01$: **, $P \leq 0.05$: *, not significant: ns.

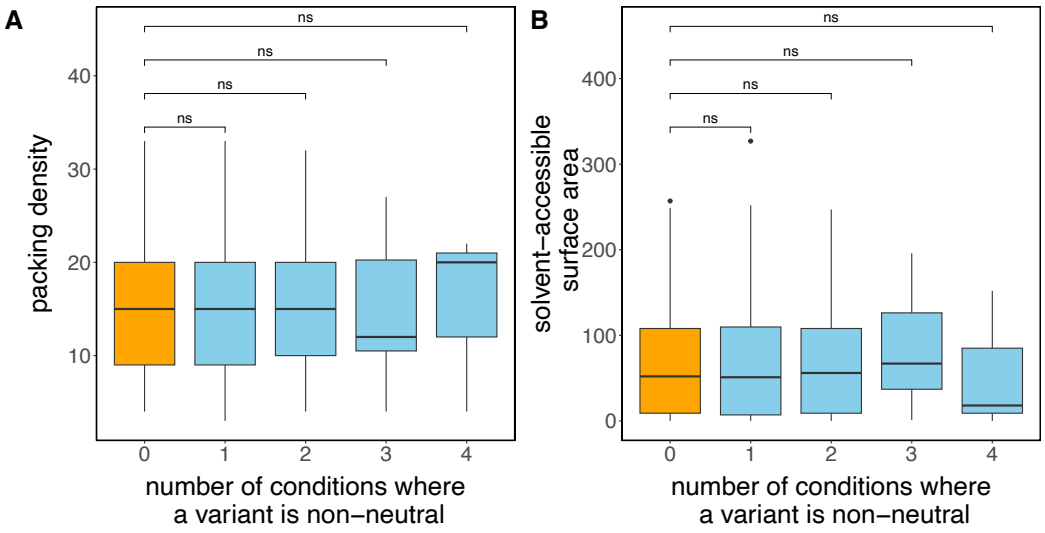


Figure S8. **Structural context does not explain synonymous fitness effects.** (A) Residue packing density and (B) solvent-accessible surface area (SASA) at codons harboring synonymous variants, stratified by the number of conditions in which each variant is non-neutral (0–4). Orange boxes indicate variants neutral in all conditions (0); blue boxes indicate variants non-neutral in at least one condition. Brackets show pairwise Wilcoxon rank-sum tests: “ns” denotes $P > 0.05$. Box plots show the median and interquartile range; whiskers extend to 1.5 times the interquartile range.

482 ***A substantial fraction of non-neutral variants shows high allele frequency***

483 Having established that non-neutral variants in our CRISPEY-BAR+ assay have larger
484 molecular impacts than neutral variants, we next set out to explore their allele distribution across
485 1,011 natural isolates of *S. cerevisiae* spanning the global ecological diversity of this species.
486 Nearly neutral theory predicts that common and older variants should have only weak effects on
487 fitness, while rare and young variants can be deleterious and maintained in a mutation–
488 selection balance. In contrast to this expectation, we often observe non-neutral variants across
489 all frequency classes: of the 3,673 variants that are non-neutral in at least one condition, 887
490 (24.1%) are common (present in more than 10 strains). Of these 887 common non-neutral
491 variants, 122 (13%) have a maximum absolute fitness across all four conditions that is larger
492 than twice the average (~33% per transfer) (Fig. 4A). Furthermore, we detect no enrichment of
493 rare variants among large-effect non-neutrals (Fisher's exact test, odds ratio = 0.94, $P = 0.617$)
494 and there is only a weak negative relationship between maximum absolute fitness and the
495 number of strains in which a variant is found (Spearman's $\rho = -0.052$, $P = 1.6 \times 10^{-3}$).

496 We observed that non-neutral variants with effects across multiple conditions have
497 slightly lower allele frequency (Spearman's $\rho = -0.039$, $P = 0.0168$; Fig. 4B and Fig. S9A).
498 These multi-condition non-neutral variants are also somewhat more likely to be found in the
499 heterozygous state (Spearman's $\rho = 0.052$, $P = 1.6 \times 10^{-3}$; Fig. 4C and Fig. S9B), even
500 controlling for their lower allele frequency (partial Spearman's $\rho = 0.049$, $P = 3.2 \times 10^{-3}$). This is
501 consistent with the expectation that deleterious variants are more likely to be recessive and
502 have their effects masked in the largely diploid *S. cerevisiae* (*20*).

503 It is possible that the weakness of the statistical relationship between allele frequency
504 and fitness effect is due to the fact that even low-frequency variants in the sample of 1,011
505 genomes are already sufficiently common to be largely neutral (*59*). Indeed, the class of non-
506 neutral variants present in only a single strain (singletons) tend to have slightly larger impacts

507 on fitness than non-singletons (Wilcoxon's test, $P = 2.8 \times 10^{-4}$; Fig. 4D and Fig. S9C).
508 Furthermore, we can approximate the age of singleton variants by the length of the terminal
509 branch associated with the single strain in which they are found, estimated here as the total
510 number of singleton variants present in the strain (*59*), and test for the association between the
511 age of a singleton and their fitness impacts in our measurements. Although we do not find a
512 significant correlation between maximum absolute fitness and the estimated age of a non-
513 neutral singleton variant (Spearman's $\rho = -0.022$, $P = 0.450$), we observe that compared to
514 neutral singletons, singletons having non-neutral effects in multiple conditions tend to be
515 younger (Spearman's $\rho = -0.1$, $P = 6.6 \times 10^{-4}$; Fig. 4E and Fig. S9D). For example, singletons
516 that are non-neutral in at least two conditions are disproportionately found in less divergent
517 strains which on average harbor 1,688 fewer strain-specific variants (6,015 vs. 7,703;
518 Wilcoxon's test, $P = 2.7 \times 10^{-4}$).
519 Additionally, non-neutral singletons tend to co-occur within the same strain less
520 frequently than neutral singletons as the number of non-neutral conditions in which they are
521 non-neutral increases ($P < 1 \times 10^{-4}$, bootstrap test for Spearman's $\rho < 0$, Fig. 4F and Fig. S9E).
522 For example, singletons that are non-neutral in at least two conditions co-occur 1.65-fold less
523 often than neutral singletons (bootstrap test for difference in mean co-occurrence > 0, $P < 1 \times$
524 10^{-4}). This is consistent with purifying selection, which continuously purges deleterious variants
525 before multiple can accumulate within a single genome.
526 These analyses suggest the existence of a subclass of rare functional variants
527 maintained at low frequency in the population under mutation-selection balance, consistent with
528 classical, nearly neutral theory. However, despite these detectable albeit weak signatures of
529 purifying selection, we also find clear evidence that many non-neutral variants persist at high
530 frequencies. Thus, while rare variants are modestly enriched for stronger fitness effects in

531 laboratory conditions, a substantial fraction of common polymorphisms is functionally impactful,
532 and at times strongly so.
533
534
535
536
537
538
539
540
541

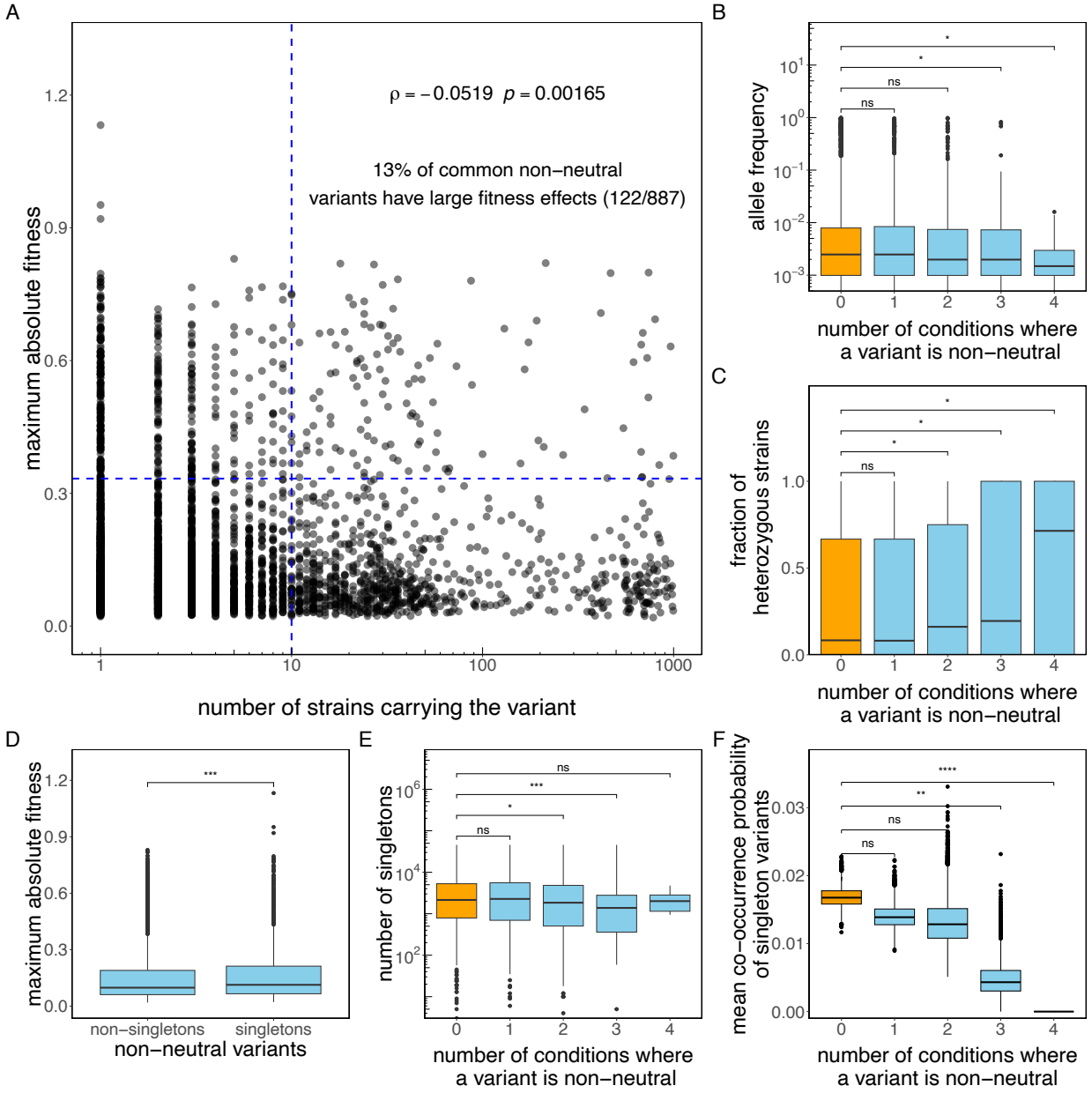


Figure 4. **A substantial fraction of non-neutral variants shows high frequency, while low-frequency variants show evidence of purifying selection.** (A) Scatter plot displaying the maximum absolute fitness values across the four conditions for non-neutral variants, plotted against the number of strains in the 1,011-yeast strain collection carrying the variant. The dashed line indicates a fitness threshold set at twice the average value (0.33). (B) Box plot showing that non-neutral variants tend to have lower allele frequency. (C) Box plot showing that the fraction of strains where the variant is in heterozygosis is higher

549 for non-neutral variants. (D) Box plot showing that the experimental fitness impact of rare variants is
550 greater than for less rare variants. Rare variants are defined as those present in a single strain
551 (singletons). (E) Box plot showing that singleton non-neutral variants tend to be younger. (F) Box plot
552 showing the co-occurrence probability for neutral and non-neutral singleton variants. Non-neutral variants
553 are grouped by the number of conditions in which they exhibit non-neutral fitness effects. The data shown
554 represent the mean co-occurrence for 10,000 bootstrap samples (Methods). Box plots show the median
555 and upper and lower quartiles; whiskers show 1.5 times the interquartile range.

556

557

558

559

560

561

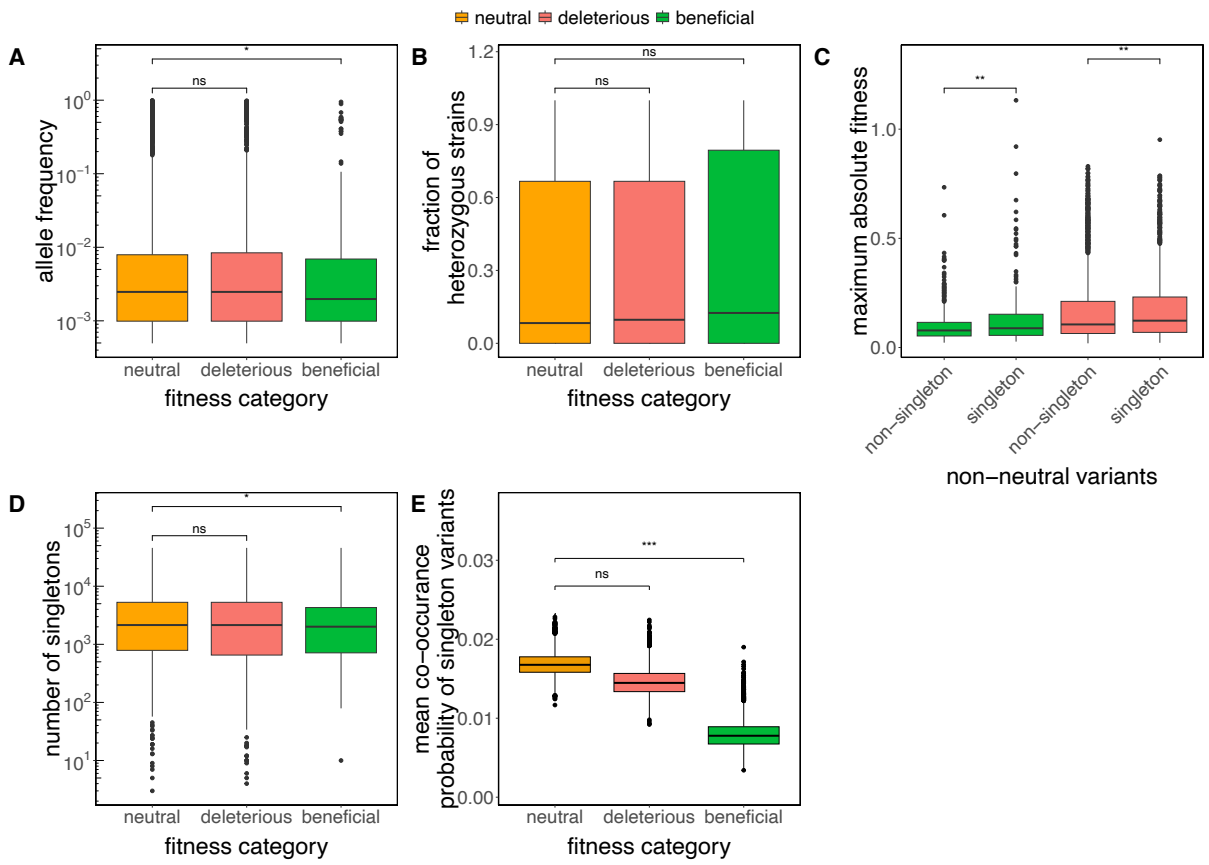


Figure S9. **To accompany Figure 4. Low-frequency variants show evidence of purifying selection.** (A) Box plot of allele frequency for neutral, beneficial, and deleterious variants. In this and all subsequent panels, non-neutral variants are grouped as beneficial if they have a positive fitness effect in at least one condition, or as deleterious otherwise. (B) Box plot showing the fraction of strains where the variant is in heterozygosis for neutral, beneficial, and deleterious variants. (C) Box plot showing that the experimental fitness impact of rare variants is greater than for more common variants for both beneficial (in green) and deleterious variants (in red). Rare variants are defined as those present in a single strain (singletons). (D) Box plot of the age of a variant for neutral, beneficial, and deleterious variants. (E) Box plot showing the co-occurrence probability of singleton variants by fitness category. The data shown represents the mean co-occurrence for 10,000 bootstrap samples (Methods). Significance levels are indicated as follows: $P \leq 0.0001$: ****, $P \leq 0.001$: ***, $P \leq 0.01$: **, $P \leq 0.05$: *, not significant: ns. Differences between groups were assessed using a Wilcoxon's test. Box plots show the median an interquartile range; whiskers extend to 1.5 times the interquartile range.

576 ***Beneficial variants show evidence of local adaptation***

577 The discovery that common variants can have very large fitness effects in laboratory conditions
578 is consistent with balance theory, in which these variants are driven to high frequencies by local
579 adaptation. However, the fitness effect of variants in experimental conditions may not reflect the
580 fitness effect in more complex natural conditions. To provide evidence for local adaptation, it is
581 essential to associate fitness effects in our assay with ecological variables. We therefore next
582 set out to explore whether non-neutral variants are more commonly found in specific clades or
583 associated with specific ecological niches.

584 *S. cerevisiae* is the dominant species in the fermentation of various beverages and foods
585 and has been domesticated on multiple independent occasions (*60*, *61*). The process of
586 domestication has had a large impact on the stress tolerance and the fermentative growth
587 capacity of this species, which are traits regulated by the core nutrient-sensing pathways we
588 study here (*62*–*64*). We therefore hypothesized that natural genetic variation in these pathways
589 may have contributed to adaptation in industrial environments.

590 We first focused on the 2,851 variants we assayed that are found exclusively in strains
591 from industrial environments (industrial variants), and compared them to the 3,250 variants
592 solely found in natural strains (Methods). As expected, industrial variants were modestly
593 enriched for non-neutral effects (Fisher's exact test, odds ratio = 1.125, $P = 0.025$). This
594 enrichment became substantially stronger when focusing on variants that are beneficial in at
595 least one condition (odds ratio = 1.56, $P = 9 \times 10^{-6}$). It disappeared when examining non-neutral
596 variants that are exclusively deleterious (Fisher's exact test, odds ratio = 0.99, $P = 0.91$) (Fig.
597 5A). These findings suggest that industrial strains are not merely enriched for functional variants
598 in general, but specifically for variants which are beneficial under the laboratory conditions we
599 tested which have previously been shown to select for Ras/PKA and TOR/Sch9 variants
600 strongly impacting fermentative performance, respiration and starvation tolerance (*46*).

601 Disentangling the role of ecology from shared ancestry is challenging because of the
602 phylogenetic structure of the 1,011-strain panel (Fig. 2A): closely related strains often share
603 both habitats and variants. To address this, we examined variants restricted to individual clades
604 and related their beneficial effects to the proportion of industrial strains within each clade.
605 Variants in clades with more industrial strains were significantly more likely to be beneficial
606 (Spearman's $\rho = 0.666$, $P = 3 \times 10^{-5}$) (Fig. 5B). To test whether this association held within
607 clades, we leveraged the fact that many clades contain both industrial and natural strains. We
608 used a logistic mixed-effects model to quantify the effect of a strain having an industrial origin on
609 the likelihood that a variant is beneficial with the clade identity as a random effect (Methods).
610 Within clades, variants found exclusively in industrial strains are significantly more likely to be
611 beneficial than those found exclusively in natural strains (effect size = 0.43, $P = 8.8 \times 10^{-4}$). Our
612 model shows that the observed enrichment of beneficial variants persists even after controlling
613 for shared evolutionary history, suggesting local adaptation within industrial environments.
614 Furthermore, the enrichment of beneficial variants in industrial strains is significantly larger than
615 in null models, which control for systematic variation in allele frequency and SIFT scores
616 associated with industrial variants along the phylogeny (Methods; S10A-B and Fig. S11A-F).
617 One possible confounder is that domestication may have relaxed purifying selection via
618 population bottlenecking, allowing deleterious variants to rise in frequency in industrial strains
619 irrespective of whether they are in the Ras/PKA or TOR/Sch9 pathways (*65*). Indeed, consistent
620 with this genome-wide bottlenecking and relaxation of purifying selection, nonsynonymous
621 variants exclusive to industrial strains tend to have lower allele frequency and SIFT scores than
622 variants exclusive to natural strains, regardless of whether they are in the Ras/PKA or
623 TOR/Sch9 pathways (Fig. S10C-D). To account for this, we generated 10,000 matched control
624 sets of variants outside the signaling pathways we studied, with equivalent allele frequency and
625 SIFT score distributions to the non-synonymous variants in our Ras/PKA and TOR/Sch9

626 CRISPEY-BAR+ screen (Fig. S11G-I). For each control set, we quantified the enrichment of
627 beneficial variants as assessed by the odds ratio (Fig. 5A), clade-level correlation (Fig. 5B), and
628 the mixed-effects model (Methods). These controls failed to reproduce the observed enrichment
629 of beneficial Ras/PKA and TOR/Sch9 CRISPEY-BAR+ variant effects, suggesting that the
630 signal we detect is not simply due to bottlenecking but instead reflects Ras/PKA- and
631 TOR/Sch9-specific local adaptation to industrial environments.

632 Finally, we set out to identify the nature of the beneficial variants that are enriched in
633 industrial strains. We first asked whether these beneficial and exclusively industrial variants
634 would be confined to one of the two major pathways or to a specific subset of genes. We found
635 that the enrichment holds for variants in both the Ras/PKA and TOR/Sch9 pathways when
636 analyzed independently (Fig. S12A). These variants are found across a broad range of genes in
637 the pathways, affecting 23 of the 24 genes we have studied (Fig. 5C and Fig. S12B). Three
638 genes show a significant enrichment of such industrial and beneficial variants as assessed by
639 the odds ratio, clade-level correlation, and the mixed effect model (Fig. S12B): the adenylate
640 cyclase *CYR1*, the effector kinase *SCH9*, and *GPB1*, a negative regulator of cAMP/PKA
641 signaling (but not its paralog *GPB2*). Of these, only *SCH9* is significantly enriched for beneficial
642 variants more generally (Fig. S4B), indicating that the enrichment in industrial strains reflects
643 selection for specific alleles that are advantageous in industrial environments rather than simply
644 increased statistical power when genes carry more variants that are beneficial in our pooled
645 competition assays.

646 These results leave open the question as to whether these signatures of local adaptation
647 are unique to the Ras/PKA and TOR/Sch9 pathways or occur more broadly across the genome.
648 Across these two pathways, 3.67% of nonsynonymous variants are present in more than 10
649 strains and had a SIFT score below 0.05—a proportion significantly higher than the genome-
650 wide average of 2.6% (two-proportion Z-test, $P = 6.45 \times 10^{-6}$; Fig. S13A). This enrichment was

651 robust to the thresholds used to define common and functional variants, supporting the idea that
652 Ras/PKA and TOR/Sch9 are potential genomic hotspots of local adaptation (Fig. S13B).
653 Nevertheless, many other genes outside these pathways also showed strong enrichment for
654 common variants with predicted impact on molecular function (Fig. S13C).

655 We also find that functional variants (SIFT score < 0.05) found exclusively in industrial
656 strains are enriched in the Ras/PKA and Tor/Sch9 pathways relative to the rest of the genome,
657 whereas this is not true for exclusively natural variants, consistent with the idea that these
658 signaling pathways underpin local adaptation to industrial environments (Fig. S13D). After
659 correcting for multiple testing, we identified 26 genes with significant enrichment for functional,
660 industrial variants (Fig. S13E). Notably, several of these genes have been previously implicated
661 in industrial adaptation (*49*, *66*). Gene-set enrichment analysis suggests that these genes are
662 disproportionately involved in carbohydrate metabolism—a key trait likely advantageous in
663 nutrient-rich, stress-prone industrial fermentations (Fig. S13E). Importantly, this genome-wide
664 enrichment analysis lacks the statistical power to recover the Ras/PKA and TOR/Sch9 genes
665 enriched for beneficial and exclusively industrial variants, suggesting that many other genes
666 may be involved in local adaptation to industrial fermentation environments. This demonstrates
667 why identifying functional variants with CRISPEY-BAR+ and then testing their ecological
668 enrichment is essential for detecting genomic hotspots of local adaptation that would otherwise
669 be missed if we rely solely on predicted molecular phenotypes.

670 .

671

672

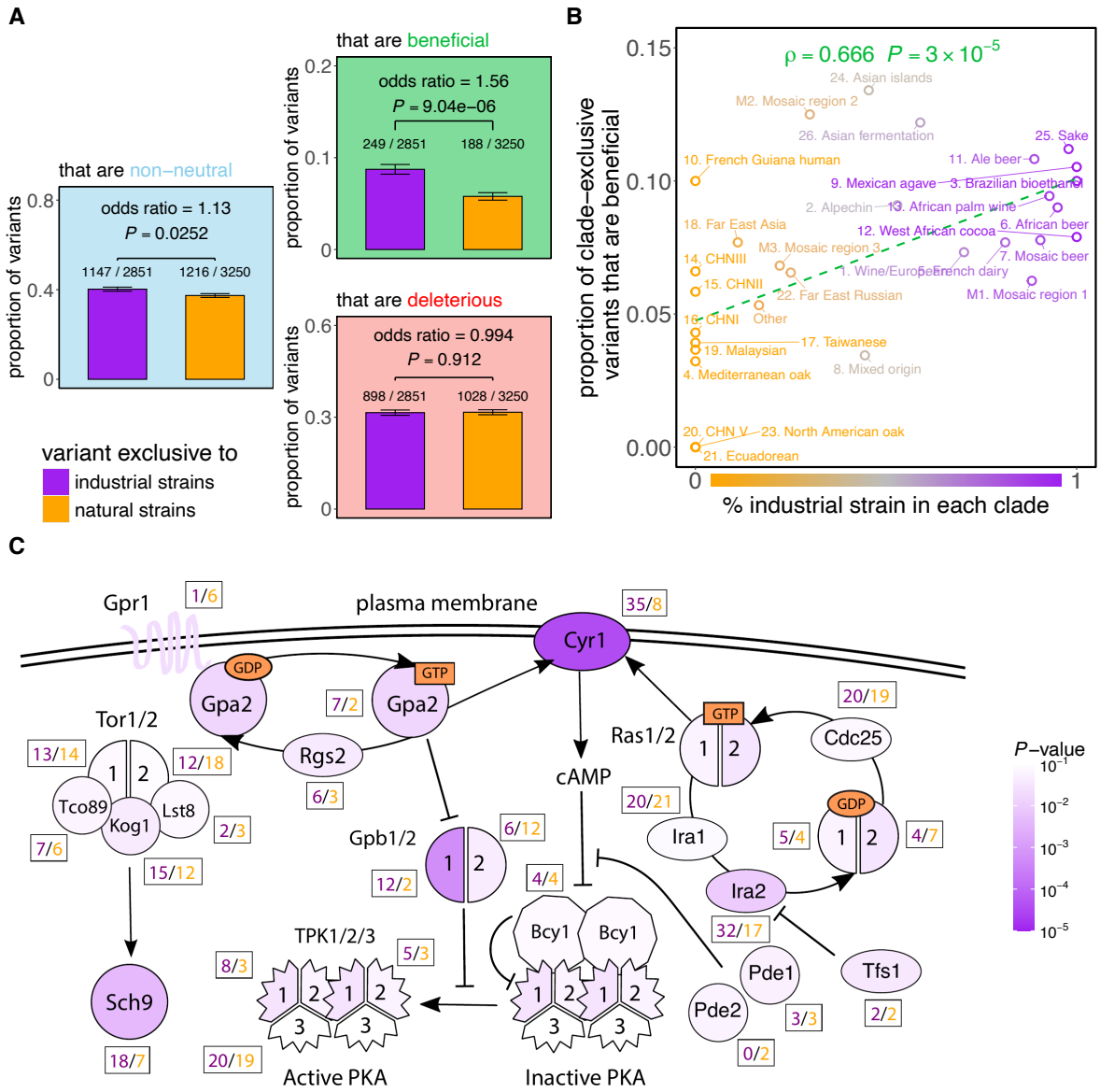


Figure 5. Beneficial variants show evidence of local adaptation. We classify variants according to whether they are exclusively observed in yeast strains from natural origins (orange) or industrial origins (purple) (Methods). (A) Bar plots comparing the proportion of variants in each group that are non-neutral (blue), beneficial (green) or deleterious (red). Each of the 26 clades shown in Figure 2A has several clade-exclusive variants, and many contain a mixture of industrial and natural strains. (B) Scatter plot displaying the proportion of clade-exclusive variants that are beneficial against the proportion of strains in each clade that are from an industrial origin. (C) Schematic of the signaling pathways Ras/PKA and

681 TOR/Sch9 in yeast. The number of beneficial variants that are exclusively industrial (purple) or natural
682 (orange) for each gene is listed. Genes are colored by the odds-ratio *P*-value (FDR-corrected) of the
683 enrichment for exclusively industrial variants (Fig. S11B).

684

685

686

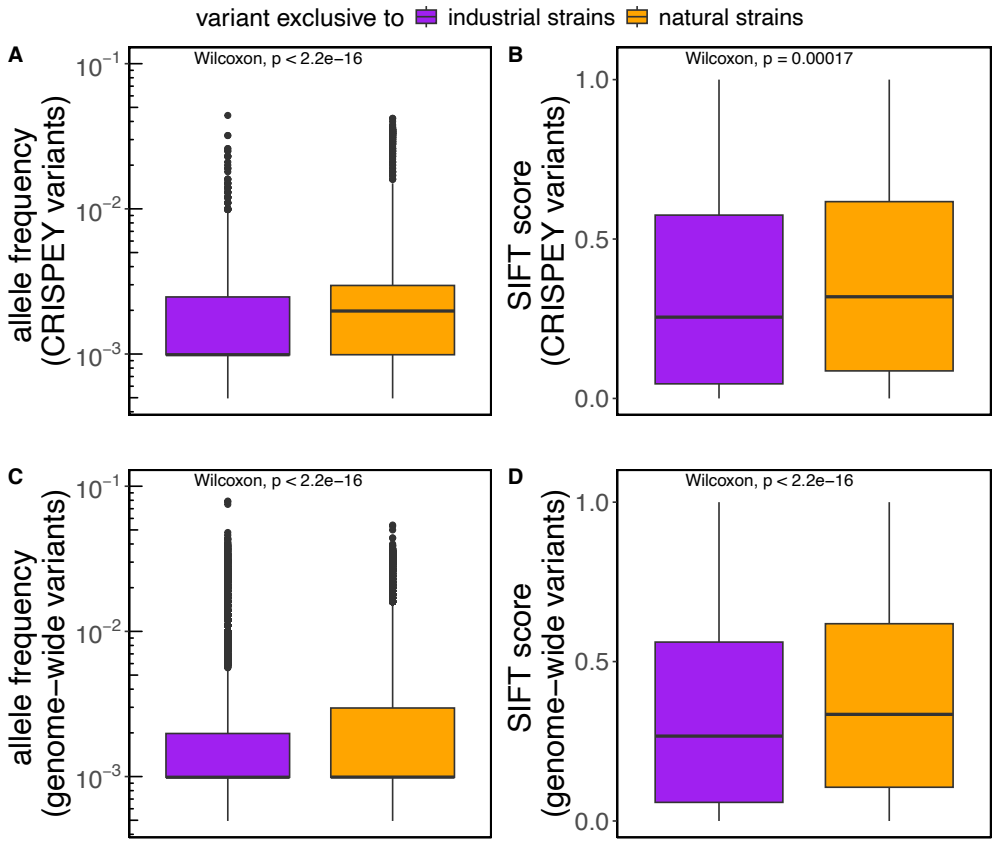
687

688

689

690

691



692

693 **Figure S10. Comparison of allele frequency and predicted functional impact (SIFT score) for**
694 **variants exclusive to industrial or natural strains.** (A-B) Box plots showing differences in allele
695 frequency (A) and SIFT score (B) for CRISPEY variants exclusively found in industrial or natural strains.
696 (C–D) Box plots showing differences in allele frequency (C) and SIFT score (F) for nonsynonymous
697 variants outside of the 24 Ras/PKA and TOR/Sch9 genes we have studied. Box plots show the median
698 and upper and lower quartiles; whiskers show 1.5 times the interquartile range. *P*-values are from
699 Wilcoxon rank-sum tests.

700

701

702

703

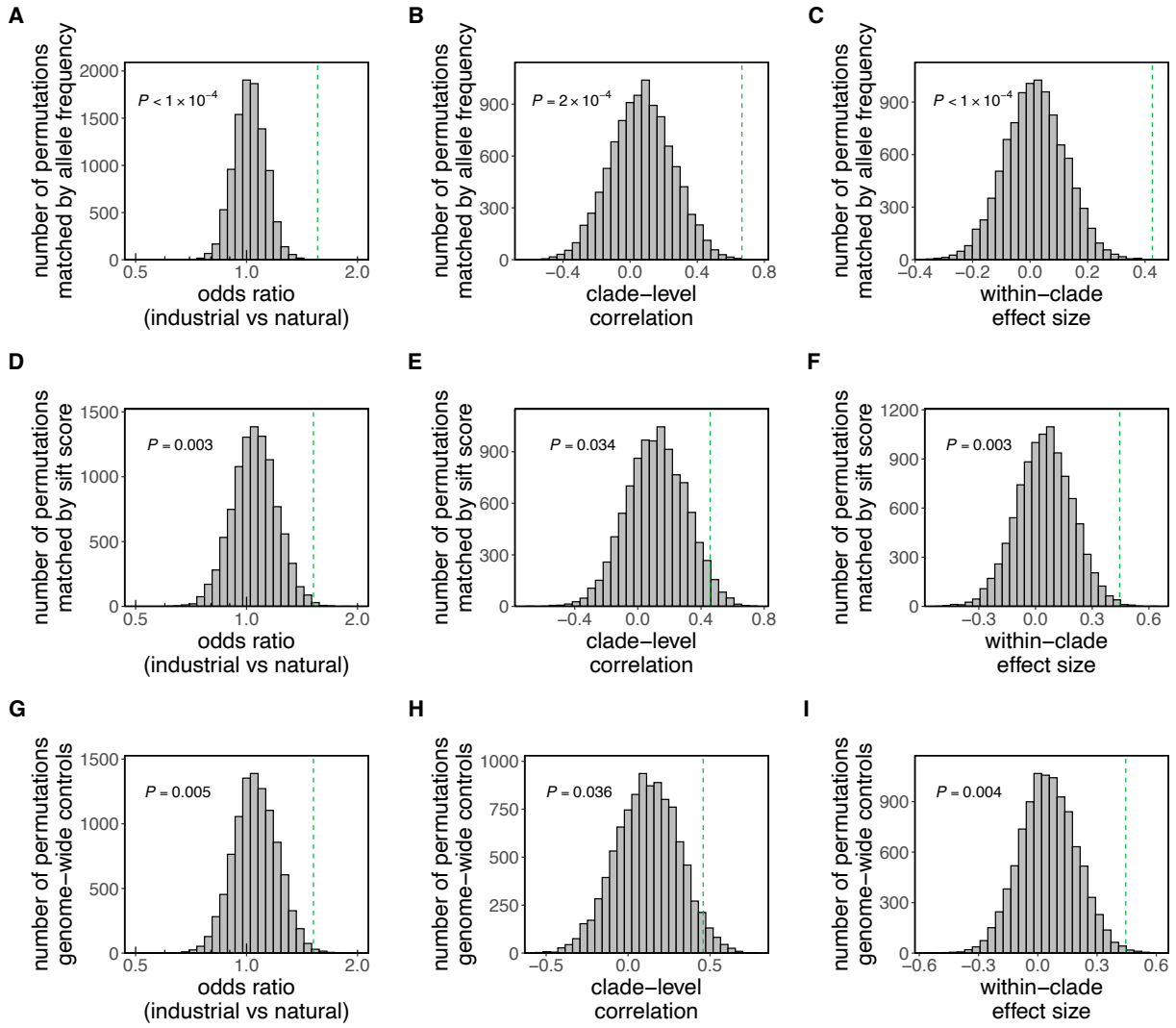
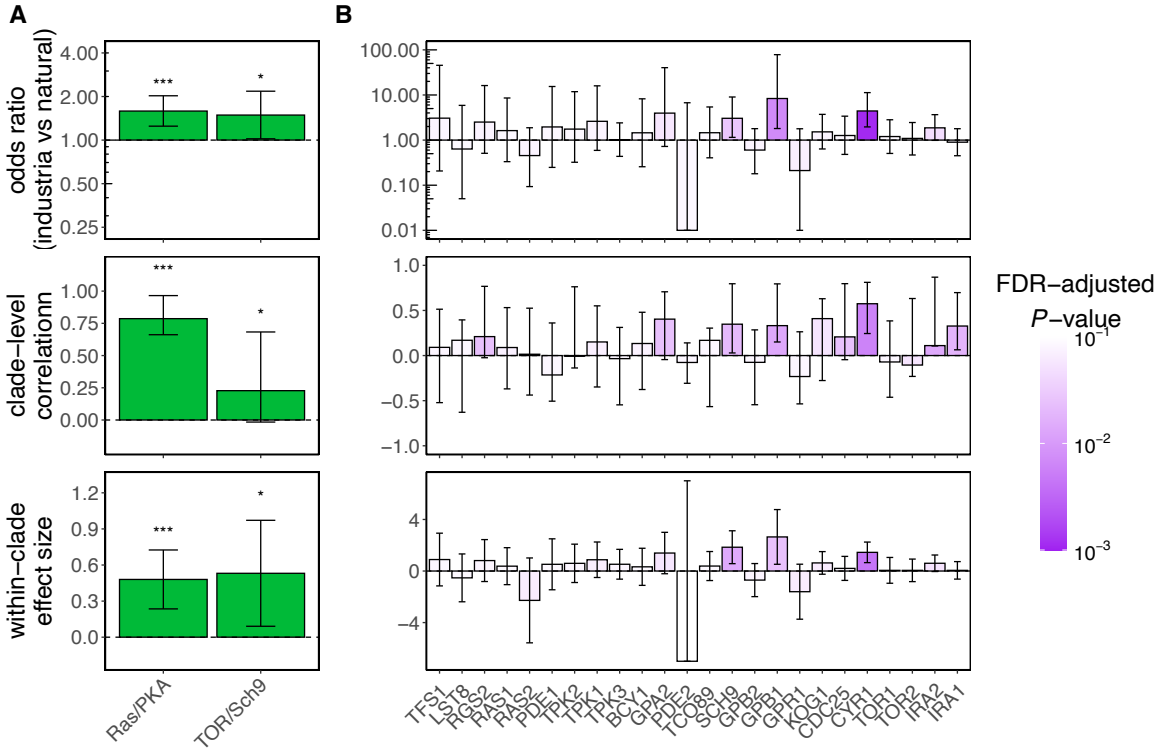


Figure S11. Permutation-based analysis of the significance of enrichment in beneficial variants in industrial strains, accounting for possible confounders. (A–C) Null distributions from allele-frequency–matched permutations of the measured variants, showing observed odds ratio (A), clade-level correlation (B), and within-clade model effect size from a mixed-effect logistic regression (Methods) (C). (D–F) Same as A–C, but with permutations matched on SIFT score for non-synonymous variants. (G–I) Same as A–C, but with genome-wide nonsynonymous variants (outside studied-targeted genes) as controls matched by allele frequency and SIFT score (Methods). Dashed vertical lines indicate observed enrichment; *P*-values reflect the fraction of permutations exceeding the observed enrichment.



714

715 **Figure S12. Pathway and gene-level analysis of beneficial variant enrichment in industrial strains.**
716 (A) Pathway-level analysis showing (from top to bottom) the odds ratio of beneficial variant enrichment in
717 industrial strains relative to natural strains, the clade-level correlation (Spearman's ρ), and the within-
718 clade effect size from a mixed-effect logistic regression (Methods). Error bars represent 95% confidence
719 intervals estimated via bootstrapping (Methods). Significance levels are indicated by asterisks (**P* < 0.05;
720 ***P* < 0.01; ****P* < 0.001; ns = not significant). The horizontal dashed line indicates the null expectation
721 (odds ratio = 1 for enrichment, correlation = 0, effect size = 0). (B) Gene-level analysis showing odds
722 ratios of beneficial variant enrichment (top), clade-level correlations (middle), and within-clade effect sizes
723 (bottom). Bars are colored by log10-transformed FDR-adjusted p-values, with error bars indicating 95%
724 confidence intervals (Methods). Horizontal dashed lines indicate null expectations as in (A). Gene names
725 are ordered by increasing protein length and displayed on the x-axis with labels rotated for readability.

726

727

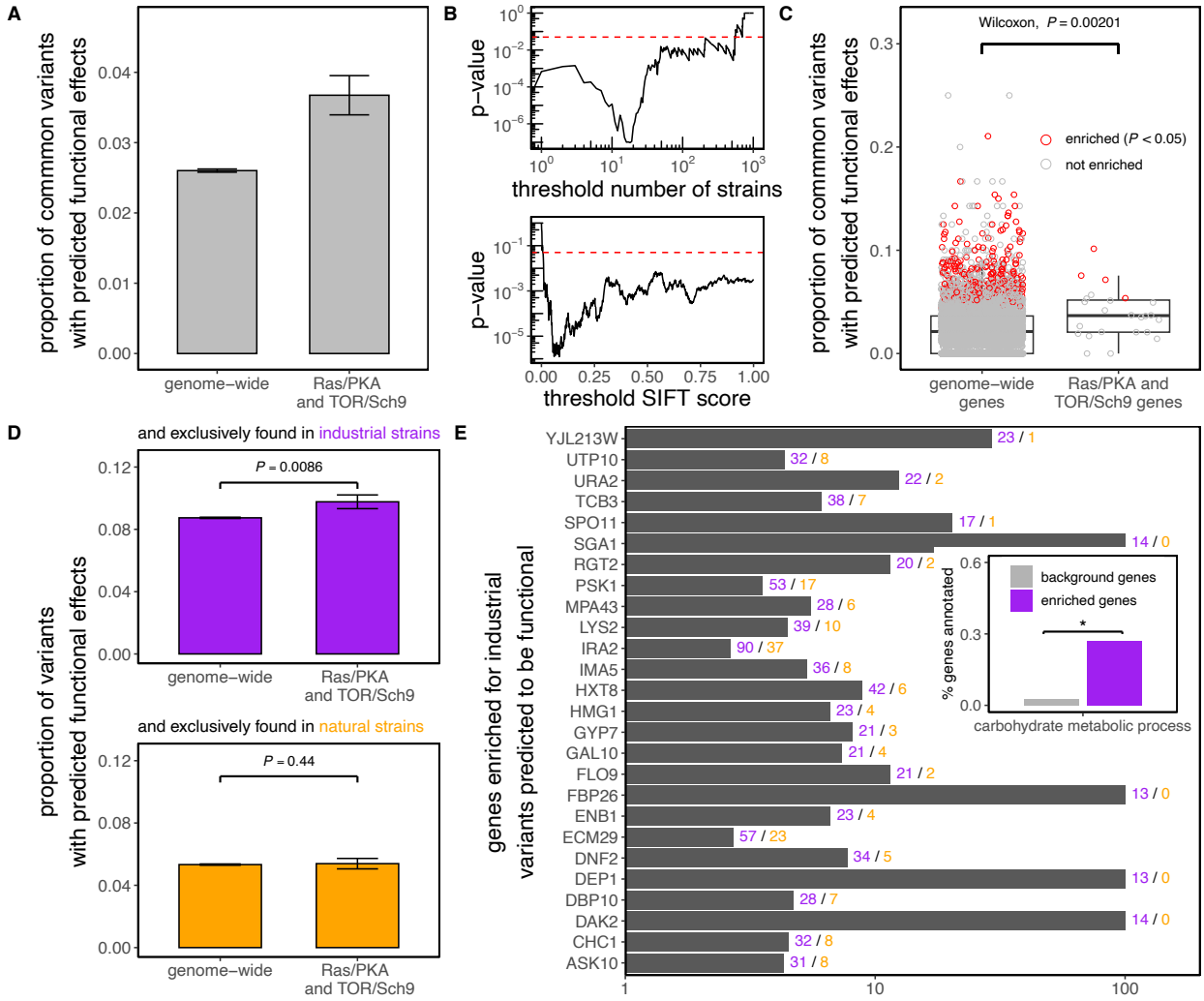


Figure S13. Ras/PKA and TOR/Sch9 genes are genomic hotspots for local adaptation to industrial environments, although other genes likely contribute. (A) Bar plot showing the proportion of non-synonymous variants that are common (>10 strains) and predicted to be functional (SIFT score < 0.05) for RAS/PKA and TOR/Sch9 pathway genes compared to genome-wide genes. Differences were assessed using Fisher's exact test. (B) The higher proportion of low-SIFT, common non-synonymous variants among RAS/PKA and TOR/Sch9 genes remains robust when varying the minimum number of strains required to define a "common variant" and the maximum SIFT score used to classify a variant as functional. (C) Boxplot showing the proportion of common, functional non-synonymous variants for all 5,359 *S. cerevisiae* genes. RAS/PKA and TOR/Sch9 genes are indicated separately, and generally have higher proportions (Wilcoxon's test). Individual genes enriched relative to the genomic background are

739 highlighted in red (Fisher's exact test, $P < 0.05$). (D) Bar plots showing the proportion of non-synonymous
740 variants that are predicted to be functional and exclusive to either industrial (purple) or natural (orange)
741 strains. Variants in RAS/PKA and TOR/Sch9 genes are compared to genome-wide genes. Statistical
742 significance was assessed using Fisher's exact test ($P < 0.05$). (E) Genes significantly enriched for
743 industrial variants predicted to be functional (FDR-adjusted $P < 0.05$) are shown as a bar plot of odds
744 ratios (industrial vs. natural strains). Counts of variants in industrial and natural strains are indicated to the
745 right of each bar (purple/industrial, orange/natural). Inset shows Gene Ontology (GO) biological process
746 enrichment for the target genes. Only carbohydrate metabolic process is significantly enriched, indicated
747 by an asterisk (*). Bars show the proportion of target genes annotated to this term compared to
748 background genes. In panels A and D bars represent mean ± standard error.
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765

766 **Discussion**

767 This study provides a comprehensive interrogation of the fitness consequences of natural
768 genetic variation within the Ras/PKA and TOR/Sch9 pathways in *S. cerevisiae*, which are
769 ancient and key regulators of cell proliferation across all eukaryotes. Leveraging our high-
770 throughput CRISPEY-BAR+ approach, we precisely measured the fitness effects for over 9,000
771 natural variants across multiple conditions, providing insights into the evolutionary forces
772 shaping these core pathways. The scale and precision of this experimental undertaking allowed
773 us to tackle a central and longstanding question in population genetics: the fitness
774 consequences of common variants in natural populations.

775 Some of our findings do agree with the expectations of nearly neutral theory. Namely,
776 some of the variants that exhibit fitness effects in our experiments show signs of negative
777 selection, as they tend to be rarer and younger, and more often found in heterozygous states,
778 which allows for the buffering of their detrimental effects on fitness, and tend not to co-occur
779 with each other (Fig. 4). These observations align with the idea that standing variation in these
780 highly conserved pathways is under purifying selection, preventing the accumulation of
781 deleterious mutations and retaining those that occur at low frequencies (*67*).

782 The majority of natural variants are neutral in our experimental conditions (61%). While
783 there may be false negatives among these variants, the high overall editing rate in our screen
784 makes it likely that most of them are indeed neutral within the limits of our detection power (~2%
785 per growth cycle). This result also follows the expectation of nearly neutral theory, particularly
786 for core eukaryotic pathways that are highly conserved from yeast to humans and are therefore
787 expected to evolve under high constraint (*68*). Nonetheless, it is remarkable that even within
788 these highly conserved pathways, nearly 40% of variants exhibit detectable and often
789 substantial functional consequences in at least one of the few environments we examined (Fig.
790 2).

791 Furthermore, we observed a substantial set of common (present in more than 10 strains
792 and often at a much higher frequency) non-neutral variants with high fitness in at least one of
793 the environments we explored (Fig. 2 and Fig. 4A). This does not naturally follow the nearly
794 neutral expectations and shows that the high frequency of a polymorphism cannot be used to
795 infer that it has a very weak functional impact, despite common practice (*69*–*72*). Moreover, we
796 find that those variants that have beneficial fitness effects in at least one of the conditions we
797 used, are enriched in strains associated with human-made environments, suggesting they may
798 have contributed to local adaptation to specialized ecological niches (Fig. 5). The process of
799 domestication has significantly influenced the stress tolerance and fermentative growth capacity
800 of *S. cerevisiae*, which are precisely the traits affected by the Ras/PKA and TOR/Sch9
801 pathways (*73*). This aligns well with the idea that local positive selection has affected variation in
802 these nutrient-sensing pathways. Another reason for the strong signal of local adaptation we
803 uncover may be that the relatively stable and controlled environments experienced by
804 domesticated yeast strains offer less ecological complexity compared to wild environments,
805 where mutations that lock signaling pathways into specific functional configurations could be
806 highly disadvantageous across a broader range of experienced conditions (*60*).
807 Although laboratory conditions inevitably simplify the environments experienced by yeast
808 in nature, some conditions can capture key components of natural environments—such as
809 nutrient-rich, fast-growth conditions (*74*). It is therefore not unreasonable to expect some
810 correlation between fitness measured in the laboratory and fitness in the wild. However,
811 because natural environments fluctuate, the extent to which a variant's experimental fitness
812 reflects its effect in nature will depend on how frequently each lineage encounters laboratory-
813 like conditions (*75*). This helps explain why beneficial variants are enriched in industrial strains:
814 these strains likely experience conditions that are, on average, more similar to our experimental
815 settings (*76*). Moreover, because adaptation can occur rapidly, there are likely ecological

816 “pockets” where laboratory conditions provide a meaningful proxy for natural selection in the
817 wild. If evolution were much slower, fitness effects would be averaged across many
818 environments, and capturing them would require far more complex experimental designs.
819 The Ras/PKA and TOR/Sch9 pathways are also key targets of adaptation in laboratory
820 experiments (*39*, *42*, *46*), suggesting they may have unique characteristics not shared by other
821 signaling pathways that makes them particularly prone to accumulating local beneficial
822 mutations. However, the signatures of domestication and local adaptation that we unveil in
823 these two signaling pathways are unlikely to be confined to them. Indeed, we find numerous
824 other genes with strong enrichments for exclusively industrial variants predicted to be functional
825 (Fig. S13). Many of these are involved in carbohydrate metabolism, which is a function likely
826 advantageous in nutrient-rich, stress-prone industrial fermentations. Notably, several of these
827 genes have been previously implicated in adaptation to industrial environments (*49*, *66*),
828 underscoring that domestication in *S. cerevisiae* is polygenic and shaped by multiple adaptive
829 routes. Therefore, applying our approach to additional pathways could reveal whether these
830 systems influence adaptation in distinct ways.
831 Our results overall agree more strongly with balance theory than with nearly neutral
832 theory. While most variants in these pathways are neutral in our lab environments or subject to
833 purifying selection, we find a substantial fraction of common variation under local positive and
834 overall balancing selection (*77*). With the advent of methods for fitness mapping *en masse* at
835 single-nucleotide resolution, we can begin to disentangle and quantify, for the first time, the
836 relative contributions of neutral evolution and balancing selection in shaping standing genetic
837 variation.
838 Despite the strength of our findings, this study has several limitations. First, we
839 conducted our fitness assays in a haploid background, while many yeast strains are naturally
840 diploid. Fitness effects in diploid strains could differ for variants with recessive or partially

841 dominant effects (*78*). Second, we focused on four environmental conditions, and exploring a
842 broader range of conditions could reveal additional insights into the pleiotropic effects of these
843 variants (*29*). Third, fitness effects smaller than our lower limit of detection (~2% per transfer)
844 could still be highly impactful in the large populations often found in both natural and industrial
845 environments. Finally, by measuring the fitness of individual variants in a single genetic
846 background, we could not investigate the potential role of epistasis in modulating variant fitness
847 across different genetic contexts and environments (*79*, *30*, *80*, *81*). Taken together, these four
848 limitations suggest that the true number of variants with functional consequences may in fact be
849 greater than what we estimate here.

850 Future studies using our method could address these limitations by measuring fitness in
851 a wider range of strains, including diploid strains, under a broader range of environmental
852 conditions, such as more complex media that mimic natural ecological conditions. However,
853 despite all the complexities that we did not explore in this study, one of its more surprising
854 outcomes is the strong correlation between fitness effects measured in the laboratory and
855 ecological adaptations observed in nature. Our work demonstrates that laboratory-based fitness
856 assays can be relevant to real-world evolutionary outcomes. Looking ahead, integrating
857 massively parallel fitness measurements with modern machine learning approaches offers
858 exciting opportunities to build more accurate, mechanistic, and predictive models of how
859 genomic information, at nucleotide resolution, gives rise to phenotypes and fitness (*82*).

860 In conclusion, this study bridges the gap between population genetics and functional
861 genomics by providing direct measurements of fitness effects for natural variants across entire
862 pathways. We constructed a comprehensive pathway-wide, nucleotide-level fitness landscape
863 of natural variation, demonstrating how massively parallel fitness mapping methods can be used
864 to address long-standing questions in evolutionary biology by leveraging ecological,
865 phylogenetic, and allele-frequency information. Our findings are more consistent with

predictions of balance theory than with nearly neutral theory, highlighting how natural selection can actively maintain functional genetic diversity in key signaling pathways. Future studies that expand this approach to additional pathways, environments, and genetic backgrounds will further our understanding of the genotype-fitness map, providing a more nuanced view of the evolutionary dynamics that shape natural populations.

Acknowledgments

This work was supported by the Swiss National Science Foundation (P2ZHP3_174735 to J.A.-R.), the European Molecular Biology Organization (ALTF 724-2018 to J.A.-R.), the CZ Biohub (to D.A.P.), the National Institutes of Health (5R35GM118165-07 to D.A.P.; GM156526 to H.B.F.; DP2-GM119140, RF1-AG057334, R01-AG06341801, and R01-HG012366 to D.F.J.), the National Science Foundation (NSF-MCB116762 to D.F.J.), a Searle Scholar Award (14-SSP-210 to D.F.J.), a Kimmel Scholar Award (SKF-15-154 to D.F.J.), a Vallee Scholar Award (to D.F.J.), and a Discovery Innovation Award from Stanford University (to D.F.J.). We are grateful to Roy Ang, Stefan Oliver Bassler, Shreyas Gopalakrishnan, Chris Jakobson, Bernard Kim, Grant Kinsler, Alexandra Khristich, and Alan Su, and all other members of the Petrov Lab and Jarosz Lab for discussions during experimental design, experimental work, and data analysis.

Author contributions

J.A.-R. and D.A.P. conceived the project. J.A.-R. conducted experiments, collected data, and performed analysis. D.A.P. and D.F.J. provided supervision and advised on experimental design and data analyses. S.A.C and H.B.F. provided extensive feedback on experimental design and helped with experimental troubleshooting. J.A.-R., J.C.C.V. and D.A.P. interpreted results. J.C.C.V. conducted phylogenetic analyses and analyses related to evidence for local adaptation, prepared figures, and wrote the results and methods sections for those analyses. J.A.-R., M.R.-M. and O.M.G. performed fitness estimation. M.R.-M. wrote the section about fitness estimation on the Methods. J.A.-R., D.F.J., H.B.F., and D.A.P. acquired funding. J.A.-R. wrote the original draft of the paper and prepared figures. J.A.-R., J.C.C.V., and D.A.P. reviewed and edited the paper with input from M.R.-M., O.M.G., D.F.J., and H.B.F.

915 **Data availability**

916 All sequence data have been deposited to SRA under NIH BioProject number

917 (PRJNA1238315).

918 **Code availability**

919 This repository contains the data and code necessary to reproduce the figures in this paper:

920 https://github.com/jaguilarrod/signaling_pathways/

921 **Methods**

922 ***Media***

923 We performed competition experiments in M3 media (*51*). This medium is glucose-limited (1.5%

924 glucose), meaning that the cells exhaust their glucose supply before any other nutrient. This

925 medium has been widely used in experimental evolution before (*83*, *42*, *46*, *84–88*). We

926 supplemented M3 with uracil (20 mg/L), L-leucine (100 mg/L), and L-lysine HCl (30 mg/L) to

927 account for the auxotrophies of ZRS111 (M3+uracil+lysine+leucine). We used SD-histidine-

928 uracil 2% raffinose media for pre-culturing before induction: 6.7 g/L Yeast Nitrogen Base (YNB)

929 (RPI), 1.92 g/L dropout mix synthetic minus histidine, uracil, adenine rich without YNB (US

930 Biological), and 20 g/L raffinose (Sigma). We used SD-histidine-uracil 2% galactose media for

931 induction, consisting of 6.7 g/L YNB, 1.92 g/L dropout mix synthetic minus histidine, uracil, and

932 adenine rich without YNB, and 20 g/L galactose (Sigma). We also used SC media: 1.7 g/L YNB;

933 5 g/L ammonium sulfate (ACROS Organics), 1.9 g/L dropout synthetic mix complete without

934 nitrogen base (US Biological), and 20 g/L glucose (Sigma).

935 ***Variant selection and editing library design***

936 We selected natural genetic variants from the 1,011 yeast genomes project (*20*). We selected

937 24 genes: 18 from the Ras/PKA pathway and 6 from the TOR/Sch9 pathway (Supplementary

938 Table S1). For each gene, we selected variants located within the coding region as well as

939 those within 500 base pairs upstream and downstream of the coding region, therefore including
940 not only coding variants but also noncoding variants that could potentially impact regulation and
941 gene expression. We designed CRISPEY oligos to edit these variants in the ZRS111 strain,
942 which contains the S288c reference alleles. The library includes 19,906 guide/donor
943 oligonucleotides targeting 11,058 genetic differences. We designed the guides and donor
944 sequences in the oligos as described previously (*26*). We also included 50 pairs of oligos
945 targeting essential genes. In each pair, one oligo introduces a nonsense mutation to a cysteine
946 or a tyrosine codon, while the other oligo introduces a synonymous mutation in the same codon.
947 Lastly, we designed 30 guides and donors with random sequences, which are therefore non-
948 editing (Fig. 1D).

949 ***Library cloning***

950 We ordered an oligonucleotide library from Twist Biosciences. We amplified the oligonucleotide
951 library using primers #310 and #313 with Q5 hot-start DNA polymerase (New England Biolabs).
952 All the primer sequences used in this study are listed in Chen *et al.* (*29*). This PCR reaction
953 produced amplicons with flanking 20 bp sequences homologous to plasmid pSAC119. We
954 cloned the PCR-amplified double-stranded DNA into NotI-digested pSAC119 using NEBuilder
955 HiFi DNA Assembly Cloning Kit (New England Biolabs). We performed the Gibson assembly
956 with a molar ratio of vector to insert of 1:5.

957 We transformed the assembled plasmid libraries via electroporation into Endura
958 electrocompetent cells (Lucigen). We performed five electroporations, each with 25 μL of
959 Endura cells and 80 ng of library plasmid DNA. We conducted the electroporations in 0.1cm-gap
960 cuvettes using a GenePulser (Bio-Rad). We allowed the transformed bacteria to recover in
961 recovery media (supplied with Endura competent cells) for 1 hour and then plated them on
962 LB/Carbenicillin plates for incubation at 37℃ overnight. We also plated dilutions to estimate the
963 transformation efficiency (cfu/transformation).

964 We extracted plasmids using the EZNA Plasmid Maxi Kit (Omega Biotek). We eluted the
965 plasmids in 3mL of kit-provided elution buffer. We subsequently digested the samples with NotI
966 and CIP (New England Biolabs) to linearize the empty vectors. We then ethanol-precipitated the
967 plasmids to concentrate the libraries at approximately 0.8 μg/μL.
968 The base yeast strain ZRS111 has been described previously (*26*). We performed the
969 transformation into yeast by electroporating the plasmid library into the strain ZRS111. We
970 prepared electrocompetent yeast by modifying previously described methods (*89*). We selected
971 the yeast transformations by plating on SD-HIS-URA 2% agar plates. We scraped the yeast
972 transformations off the selective plates using cell spreaders, resuspended them in SD-HIS-URA
973 and froze them in 15% glycerol stocks at -80℃ for subsequent experiments. We generated 19
974 million transformants, with an average representation of 963 cells per guide/donor
975 oligonucleotide.

976 ***Generation of the oligo-barcode map***

977 We amplified 120 ng of the maxiprep of the final plasmid library in a 400 μL Q5 polymerase
978 (New England Biolabs) reaction with 1 μM forward primer #73 and 1 μM reverse primer
979 equimolar mix of primers #327-#334. We performed the PCR reaction with 1 M betaine and an
980 initial denaturation of 98℃ for 2 min, followed by 16 cycles of 98℃ for 10 s, 65℃ for 20 s, and
981 72℃ for 20 s, and then a final extension at 72℃ for 2 min. The staggered primers provide
982 additional sequence complexity for read 2 during sequencing. We purified 400 μL of the first-
983 step PCR with the Zymo DNA Clean & Concentrator and eluted it into 30 μL of elution buffer.
984 We further purified it using a SizeSelect II gel (Invitrogen) to obtain a ~500 bp product followed
985 by purification and elution using a Zymo DNA Clean & Concentrator column in 20 μL of elution
986 buffer. The purified amplicon was indexed by Q5 polymerase (New England Biolabs) in a 500
987 μL reaction and a 1 μM equimolar mix of indexing primers for Illumina sequencing. The PCR
988 thermocycling conditions were as follows: initial denaturation at 98℃ for 2 min, followed by 5

989 cycles of 98℃ for 10 s, 70℃ for 20 s, and then a final extension at 72℃ for 2 min. We purified
990 100 µL of this second PCR using the Zymo DNA Clean & Concentrator and eluted it into 30 µL
991 of elution buffer. We further purified it using a SizeSelect II gel (Invitrogen), followed by
992 purification and elution in a Zymo DNA Clean & Concentrator column in 20 µL of elution buffer.
993 We quantified them using Qubit dsDNA HS (Invitrogen). We submitted the library to Admera for
994 paired-end sequencing on a lane of the HiSeq 4000 using a custom read1 primer (#75). We
995 used a custom script to associate barcodes with oligos from the read data.

996 ***Pooled editing and plasmid curing***

997 We carried out pooled editing for CRISPEY-BAR+, like previous work (*26*, *29*, *30*). We thawed
998 the yeast library from frozen glycerol stocks and added it to 200 mL SD-histidine-uracil 2%
999 raffinose, starting at an OD_{600} of 0.4, and incubated at 30℃ with shaking at 250 RPM for 24 h.
1000 After this phase of growth in raffinose, we induced editing with galactose. To initiate this new
1001 phase, we re-inoculated the pre-editing cultures grown in raffinose in 200 mL SD-uracil 2%
1002 galactose, with a starting OD_{600} of 0.4, and incubated at 30℃ with shaking at 250 RPM for 24 h.
1003 This phase of editing induced by galactose lasted for 72 h. We transferred the yeast library to
1004 200 mL of fresh SD-histidine-uracil 2% galactose every 24 hours. We harvested the cells from
1005 the last growth in galactose and stored them in SD-histidine-uracil 2% glucose with 15%
1006 glycerol at -80℃.

1007 We then removed CRISPEY-BAR+ plasmids after editing by growing them in 200 mL of
1008 SD 2% glucose media, starting at an OD_{600} of 0.4, and incubating at 30℃ with shaking at 250
1009 RPM for 24 h to remove plasmid selection. After this phase of non-selective growth, we
1010 inoculated 200 mL SD 2% glucose with 1g/L of 5-fluoroorotic acid monohydrate (5-FOA)
1011 (GoldBio), starting at an OD_{600} of 0.4, and incubated at 30℃ with shaking at 250 RPM for 24 h to
1012 select against the *URA*-bearing CRISPEY-BAR+ plasmid (*90*). We froze the edited, plasmid-

1013 free yeast library and stored it in 15% glycerol stocks at -80℃ in preparation for the growth
1014 competitions.

1015 ***Pooled growth competitions***

1016 We carried out pooled competitions in 500 mL DeLong flasks (with flat bottoms) in 100 mL of
1017 M3+uracil+lysine+leucine media. We thawed edited and plasmid-cured yeast cells (see *Pooled*
1018 *editing and plasmid curing*) in 15 mL of SC media at 30℃, shaking at 250 RPM overnight in a
1019 500 mL flask. We inoculated 100 mL M -uracil-lysine-leucine media starting at an OD_{600} = 0.4
1020 and grew at 30℃ shaking at 250 RPM for 24h. We then inoculated 500 µL of this saturated
1021 culture into 100 mL of fresh media in 500 mL DeLong flasks. This culture was then grown at
1022 30℃ in an incubator sharing at 250 RPM, with incubation time varying depending on the
1023 experimental condition: 24 h (1 day), 48 h (2 days), 72 h (3 days), or 120 h (5 days). We
1024 included three replicates per condition. After incubation, we transferred 500 µL of the saturated
1025 culture into fresh media in a new flask. This serial dilution was continued six times, yielding six
1026 time-points over which to measure the rate at which the barcode frequency changed. We pre-
1027 warmed fresh media before each transfer. Contamination caused the loss of the last three time-
1028 points for one of the three replicates of the 2-day condition. After each transfer of 500 µL, we
1029 froze the remaining 9500 µL to sequence the barcodes present at each time-point later. To
1030 prepare this culture for freezing, we transferred it to 50 mL conical tubes, spun it down at 3000
1031 RPM for 5 min, resuspended it in 15% glycerol, aliquoted it into three 1.5 mL tubes, and stored it
1032 at -80℃.

1033 ***DNA extractions***

1034 After completing the growth competitions, we extracted DNA from frozen samples using a
1035 modified protocol that enhances the extraction yield (*85*). For each sample, a single tube of the
1036 three that were frozen for each sample (see *Pooled growth competitions*) was removed from the

1037 freezer and thawed at room temperature. We extracted DNA from 400 µL of the sample, and
1038 stored the remaining culture at -80℃, using a modified version of the Lucigen MasterPure yeast
1039 DNA purification kit (#MPY80200). We transferred the thawed cells into a 2 mL tube and
1040 centrifuged at 3500 RPM for 3 min. After discarding the supernatant, the pellet was then
1041 resuspended in 400 µL of sterile water and spun again at 3500 RPM for 3 min. After discarding
1042 the supernatant, the pellet was resuspended with 1200 µL of the MasterPure lysis buffer
1043 supplemented with eight µL of zymolyase and four µL of RNase A (Thermo Scientific). We split
1044 the mix into two 1.5 mL centrifuge tubes, which we incubated at 37℃ for 1 hour. After such
1045 incubation, we added 2 µL of zymolyase to each tube, and then we incubated them at 65℃ for 4
1046 hours, vortexing the tubes every hour. We then placed the tubes on ice for 10 min and
1047 subsequently mixed 300 µL of MPC Protein Reagent with the solution from each tube by
1048 vortexing. We separated protein and cell debris by centrifugation at 10,000 RPM for 15 min,
1049 transferring 800 µL of supernatant from each tube to a 2 mL centrifuge tube. We further
1050 separated the remaining protein and cell debris by centrifuging at 13200 RPM for 5 min. Next,
1051 we added 800 µL of isopropanol to each tube, mixed by inversion, and then centrifuged at
1052 10,000 RPM for 15 minutes. The supernatant was discarded. We washed the pellet by adding
1053 500 µL of ethanol 70%, mixing by inversion, and then centrifuging the tubes at 10,000 RPM for
1054 7 min. We discarded the supernatant, and the tubes were left to air-dry. We resuspended the
1055 pellet, containing the DNA, in 50 µL of Elution Buffer. Next, we combined the two tubes per
1056 sample into a single 1.5 mL DNA LoBind® tube. We added 2 µL of 10 mg/mL RNAse A (Thermo
1057 Scientific) and incubated at 37℃ for 30 min to digest the RNA further, and then quantified the
1058 amount of DNA using the Qubit dsDNA BR assay (Invitrogen).

1059 ***Sequencing library preparation***

1060 We amplified 4.5 µg of genomic DNA per sample in a 1200 µL Q5 polymerase (New England
1061 Biolabs) reaction with 1 µM forward primer #261 and 1 µM reverse primer equimolar mix of

1062 primers #327-#334. We split the PCR reaction into 24 50-µL volumes and performed it with 1 M
1063 betaine and initial denaturation at 98℃ for 2 min, followed by 18 cycles of 98℃ for 10 s, 65℃ for
1064 20 s, and then extension at 70℃ for 2 min. The staggered primers provide additional sequence
1065 complexity for read2 during the sequencing of the barcode. We purified 200 µL of this first-step
1066 PCR with 200 µL of magnetic beads (NucleoMag NGS Clean-up and Size Select, REF
1067 744970.5). The purified amplicons were indexed by Q5 polymerase (New England Biolabs) in a
1068 50 µL reaction and a 1 µM equimolar mix of indexing primers for Illumina sequencing. The PCR
1069 thermocycling conditions were as follows: initial denaturation at 98℃ for 2 min, followed by 8
1070 cycles of 98℃ for 10 s, 70℃ for 20 s, then extension at 70℃ for 5 min. We purified the indexed
1071 amplicons using 50 µL of beads and eluted them in 100 µL of water. We quantified them using
1072 the Qubit dsDNA HS assay (Invitrogen). We mixed equimolar amounts of the purified, indexed
1073 amplicons from six time-point samples for each of the three replicates per condition and purified
1074 them using a SizeSelect II gel (Invitrogen) to obtain a product of ~300 bp. We then purified the
1075 size-selected libraries with beads. We submitted the libraries to the CZ Biohub SF for paired-
1076 end sequencing in a flow cell (4 lanes) of the NovaSeq S4 using a custom read1 primer (#354).

1077 ***Read processing***

1078 We processed reads in FASTQ format from the competition samples sequenced using NovaSeq
1079 using a custom pipeline. We used cutadapt (*91*) to trim adaptors from the reads. We used the
1080 following parameters for read2 trimming: 5' adaptor sequence as `GGCCAGTTTAAACTT`, 3'
1081 adaptor sequence as `GCATGGC`, maximum error rate of 0.2, and the length of the trimmed read2
1082 as 27 base pairs. We then combined trimmed paired reads using FLASH (*92*) with a minimum
1083 overlap of 12 base pairs and a maximum mismatch rate of 0.25. The resulting barcode is 27
1084 base pairs in length. The barcodes with a perfect match to those barcodes from the barcode-
1085 oligo map were counted for analysis as described below.

Estimating edited variant effect on fitness

Incorporating multiple barcodes per edit enhances the precision of fitness measurements by increasing biological replication within a flask. This approach enables the detection of random mutants that may arise during transformation, editing, or competition, as well as cases where the specific edit failed in the CRISPEY-BAR+ procedure, which appear as outliers, thereby improving the accuracy of our estimates of variant fitness effects. Spontaneous mutations with strong positive fitness effects, in particular, can disproportionately dominate the reads for a given variant. However, these mutations are unlikely to impact multiple independently edited barcodes. To mitigate false positives, we removed these outlier barcodes by discarding the barcode with the highest read count for each variant. We aggregated all the barcodes for the same edit to infer fitness from the edit count trajectories.

To infer fitness from the edit count trajectories, we followed a variational Bayesian approach (*52*). Briefly, our objective is to infer the fitness values $\vec{s} = (s^{(1)}, s^{(2)}, \dots, s^{(N)})$ for all *N* edited strains. For a specific strain *j*, our data consists of a vector of barcode counts across all time points sequenced in the experiment,

$$\overrightarrow{r^{(j)}} = (r_1^{(j)}, r_2^{(j)}, \dots, r_T^{(j)}),$$

for the *T* time points in our experiment. The entire dataset then consists of *N* vectors of raw reads

$$\underline{R} = (\overrightarrow{r^{(1)}}, \overrightarrow{r^{(2)}}, \dots, \overrightarrow{r^{(N)}}).$$

Our fitness models for how the allele frequency of barcode *j* changes over time takes the form

$$f_{t+1}^{(j)} = f_t^{(j)} \, exp\, ((s^{(j)} - \underline{s_t})\tau),$$

where $f_t^{(j)}$ is the allele frequency of barcode j at time t, $\underline{s_t}$ is the population mean fitness at time t, and τ is the time interval between t and $t + 1$. Thus, we must include the inference of the nuisance parameter $\underline{s_t}$ for each time point in our experiment. To constrain the value of this

parameter, the experiment includes a series of what we consider to be neutral barcodes (barcodes associated with non-editing oligos) whose fitness is zero by definition. This means that we include a series of barcodes that, by definition, have

$$s^{(n)} = 0.$$

Furthermore, the barcode frequency $f_t^{(j)}$ is not what our experiments quantify. Naively, we could simply assume that

$$f_t^{(j)} = \frac{r_t^{(j)}}{\sum_{k=1}^{N} r_t^{(k)}},$$

but this would assume we have perfect readouts for the experiment. Since we use a Bayesian framework, we can simply include the computation of all frequencies

$$\underline{F} = (\overrightarrow{f^{(1)}}, \overrightarrow{f^{(2)}}, \dots, \overrightarrow{f^{(N)}}),$$

as nuisance parameters in our inference.

By Bayes theorem, our inference problem can be expressed as

$$\pi(\vec{s}, \underline{\vec{s}}, \underline{F} \mid \underline{R}) \propto \pi(\underline{R} \mid \vec{s}, \underline{\vec{s}}, \underline{F})\pi(\vec{s}, \underline{\vec{s}}, \underline{F}),$$

where $\underline{\vec{s}} = (\underline{s_1}, \underline{s_2}, \dots, \underline{s_T})$, and $\pi(\cdot)$ is a probability distribution. Although this is an intractable problem, Razo-Mejia *et al.* (*52*) deploy a variational approach to solve this inverse problem of extracting fitness values from raw read counts.

Furthermore, using a hierarchical Bayesian model, the Bayesian framework can be extended to combine the information from multiple experimental replicates into a single inference problem. Hierarchical models systematically account for all available information and propagate sources of uncertainty throughout the inference process, increasing the statistical power when inferring the fitness values. The results shown in the main text take this hierarchical approach as follows: The reads for all individual barcodes mapping to a specific edit were

1133 aggregated to be used as a single number $r_t^{(j,n)}$ for edit *j* and replicate *n* at time *t*. We then
1134 assume that we can infer the corresponding fitness $s^{(j,n)}$ as a sample from a hyper-parameter
1135 distribution (the hyperfitness) $\phi^{(j)}$ shared among all edits *j* across all replicates. The objective
1136 then becomes inferring the values for all hyperfitness values $\vec{\phi} = (\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(E)})$ for all *E*
1137 edits and the individual replicate fitness values $\underline{S} = (\overrightarrow{s^{(1)}}, \overrightarrow{s^{(2)}}, \dots, \overrightarrow{s^{(M)}})$ for all *M* experimental
1138 replicates given the corresponding raw reads $\vec{R} = (\overrightarrow{r^{(1)}}, \dots, \overrightarrow{r^{(M)}})$. This inference problem can be
1139 written by Bayes theorem as (we ignore the corresponding nuisance parameters for simplicity
1140 although they are included in the full inference; see (*49*) for more details)

1141 $$\pi(\vec{\phi}, \underline{S} \mid \underline{R}) \propto \pi(R \mid \underline{S})\pi(\underline{S} \mid \vec{\phi})\pi(\vec{\phi}).$$

1142 In this way, all experiments are connected through the hyperfitness parameter, which
1143 brings together all information from each replicate and its corresponding uncertainty. The main
1144 text results are based on the posterior distributions of the hyperfitness parameters.

1145 ***Variant properties***

1146 We annotated variants using Variant Effect Predictor (VEP) (*93*). We retrieved the predicted
1147 protein structures of all *S. cerevisiae* S288C genes in Supplementary Table 1 from the
1148 AlphaFold Protein Structure Database (*94*, *95*). There are no protein structures for *IRA1*, *IRA2*,
1149 and *SDC25*. We measured solvent accessibility using DSSP (*96*, *97*). We measured packing
1150 density as the number of other α-carbons within 10Å of the α-carbon for a given residue using
1151 custom code. We used FoldX to determine the destabilizing effect of nonsynonymous variants
1152 (*58*). For all yeast genes, we retrieved SIFT scores from the mutfunc database (*98*). We
1153 obtained the allele frequency and heterozygosity of each variant from the 1,011 yeast genomes
1154 collection (*20*). From the same source, we received the number of singleton variants for each of
1155 the 1,011 strains and their ploidy. For a given singleton variant, we use the number of other
1156 singleton variants present in the same strain as a proxy for variant age, following Zhu *et al.* (*59*).

1157 While this metric serves as a reasonable proxy on average—since under a strict molecular clock
1158 with constant effective population size (N_e), branch length reflects divergence time—variation in
1159 N_e can introduce noise. Lineages with smaller N_e may accumulate more slightly deleterious
1160 mutations, leading to longer apparent branch lengths than lineages with larger N_e, even under
1161 identical divergence times.

1162 ***Constructing a maximum likelihood phylogenetic tree***

1163 To construct a maximum likelihood phylogenetic tree for the 1,011 natural isolates we used the
1164 published whole-genome sequencing data from Peters *et al.* (*20*). We started with the
1165 previously published VCF file, which contains every single-nucleotide polymorphism (SNP) and
1166 indel that has been called at the population level (1011Matrix.gcvf). We filtered for SNP sites
1167 present in all strains and where the minor allele could be confidently called in at least 2 strains
1168 and could thus be parsimony informative. A maximum likelihood tree was generated based on
1169 the 866,130 filtered SNPs using IQTREE2 (iqtree2 -s filtered_snps.min4.phy -m GTR+ASC)
1170 (*99*). The resulting maximum likelihood tree was rooted using the Taiwanese clade as an
1171 outgroup as this represented the most divergent population (*20*). This rooted tree was used in
1172 all subsequent analyses (Fig. 1B and Fig. 2A).

1173 ***Variant co-occurrence analysis***

1174 For each pair of alleles (*i* and *j*) the probability of co-occurring in the same natural isolate was
1175 calculated as the Jaccard similarity coefficient:

1176 $$J(A,B) = \frac{A \bigcap B}{A \bigcup B}$$

1177 , where *A* is the set of natural isolates in which allele *i* is present and *B* is the set of natural
1178 isolates in which allele *j* is present. Using this measure, we quantified whether pairs of non-
1179 neutral variants were significantly less likely to co-occur than pairs of neutral variants as this

1180 would suggest that they were under purifying selection (since a deleterious variant is expected
1181 to go extinct before a second deleterious variant can arise in the same genetic background).
1182 We calculated the mean Jaccard similarity for all pairs of singleton non-neutral variants
1183 and then compared this with the mean Jaccard similarity for all pairs of neutral singleton
1184 variants (Fig. 4F). To determine whether the difference in means was statistically significant, we
1185 repeated this calculation for 1,000 bootstrap replicates. The *P*-values in Figure 4F corresponds
1186 to the proportion of bootstrap replicates in which neutral variants co-occurred less frequently
1187 than the non-neutral variants considered.

1188 ***Identifying exclusively natural, exclusively industrial, and clade-exclusive variants***
1189 Of the 1,011 natural isolates reported in Peters *et al.* (*20*), 963 isolates belong to one of 29
1190 distinct clades (Fig. 2A), and were assigned to one of 23 different ecological origins in that
1191 study. We grouped these ecological origins into three categories: industrial, natural, and other
1192 (Fig. 1B). Specifically, 587 strains were classified as “industrial” if their ecological origin was
1193 listed as either beer, bakery, industrial, wine, distillery, palm wine, bioethanol, fermentation,
1194 cider, dairy or sake. We classified 392 strains as “natural” if their ecological origin was listed as
1195 either water, fruit, flower, human-clinical, insect, nature, tree, soil, or human. Finally, 32 strains
1196 were classified as “other” if their ecological origin was listed as either laboratory strains,
1197 probiotics, or unknown. Strains in this last category were excluded from subsequent analyses.
1198 Using the VCF file from Peters *et al.* (*20*), we determined the ecological and clade-level
1199 distribution of 9,447 variants from the Ras/PKA and TOR/Sch9 pathways for which we could
1200 estimate selection coefficients in our pooled competition assays. Of these variants, 2,851 were
1201 observed exclusively in industrial strains, whereas 3,250 were exclusive to natural strains. At
1202 the clade level, 5,778 variants were restricted to a single clade, whereas 2,994 were present in
1203 multiple clades. The remaining 675 variants were found in the 48 strains without clade

assignments and were treated as a single group in clade-level analyses (see *Quantifying variant enrichment in industrial strains*).

Quantifying the enrichment of beneficial variants in industrial strains

We determined whether beneficial variants were disproportionately enriched in industrial versus natural strains using three complementary statistics each capturing variant enrichment at different phylogenetic scales. First, as a straightforward measure of enrichment, we calculated the odds ratio (OR) comparing the likelihood that an exclusively industrial variant was beneficial to that of a solely natural variant (Fig. 5A). Second, we computed the Spearman's rank correlation (ρ) between the proportion of clade-exclusive variants that were beneficial and the fraction of industrial strains in each clade, to assess how enrichment in beneficial variants tracks phylogenetic structure (Fig. 5B). Third, we fit a generalized linear mixed-effects model (GLMM) with clade as a random effect to estimate the association between ecological origin and the probability of being beneficial while controlling for shared ancestry.

For the OR, statistical significance and 95% confidence intervals were obtained using the *fisher.test* function in R. For ρ, we calculated the correlation between (a) the proportion of clade-exclusive variants that were beneficial and (b) the proportion of industrial strains in each clade. To account for sampling uncertainty, we generated 10,000 bootstrap replicates by resampling clades with replacement, weighting each by the number of variants it contained. The bootstrap distribution was used to derive 95% confidence intervals and p-values. For the GLMM, we fit a binomial model of the form *IsBeneficial ~ Type_Eco + (1 | Clade)* using the *lme4* package in *R*. The fixed-effect estimate for the "industrial" category gives the log-odds difference in the probability of a variant being beneficial relative to exclusively natural variants, accounting for clade-level differences. 95% confidence intervals were computed using the *confint* function, and *P*-values were obtained from the model summary.

1228 We repeated the above analysis for deleterious non-neutral variants and never observed
1229 any statistically significant enrichment in industrial strains for any of our three statistics (results
1230 for the odds ratio are also shown in Fig. 5A). We also repeated the above analysis separately
1231 for variants found within one of the two pathways considered (Fig. S10A), and for variants in
1232 each of the 24 individual genes we have studied (Fig. S10B).
1233

1234 ***Permutation tests controlling for allele frequency and functional impact***

1235 To test whether the enrichment of beneficial Ras/PKA pathway variants in industrial strains
1236 could be explained by differences in allele frequency or predicted functional impact (SIFT
1237 score), we performed permutation analyses matched on these properties. Specifically, we
1238 binned all variants into 20 roughly equal-width bins based on log10-transformed allele
1239 frequencies, or separately into 20 bins based on their SIFT scores (for nonsynonymous variants
1240 only). Within each bin, we permuted the beneficial status of variants without replacement,
1241 thereby preserving the original distribution of allele frequencies or SIFT scores while
1242 randomizing whether or not a given variant is beneficial or not.
1243 For each of 10,000 permutations, we recalculated the three statistics described above—
1244 odds ratio, clade-level Spearman correlation, and GLMM effect size—providing a null
1245 distribution under the hypothesis that enrichment arises solely from differences in allele
1246 frequency or predicted functional impact. The observed statistics were then compared to the
1247 permutation distributions to assess whether they could be explained by these variant properties
1248 alone (Fig. S9A-F).
1249 To test whether the observed enrichment of beneficial variants in industrial strains was
1250 specific to Ras/PKA and TOR/Sch9 pathways or could be explained by genome-wide effects
1251 such as relaxed purifying selection due to domestication bottlenecks, we performed permutation
1252 tests using matched control sets of non-synonymous variants outside these pathways. We first

1253 identified non-synonymous variants exclusive to industrial or natural strains, excluding those
1254 missing SIFT scores. Variants within the Ras/PKA and TOR/Sch9 pathways ("CRISPEY"
1255 variants) were flagged, and their beneficial status was assigned based on the CRISPEY screen
1256 results.

1257 To generate genome-wide controls, we matched each non-synonymous CRISPEY
1258 variant to one or more genome-wide non-synonymous variants with identical SIFT score and
1259 log10-transformed allele frequency bins (20 bins each). For each of 10,000 permutations, we
1260 randomly sampled one matched variant per CRISPEY variant, assigning it the original beneficial
1261 status of the corresponding CRISPEY variant. We then recalculated the odds ratio, clade-level
1262 Spearman correlation, and GLMM effect size for matched controls to obtain null distributions
1263 that account for differences in allele frequency and predicted functional impact across the
1264 genome (Fig. S9G-I).

1265

1266

1267

1268

1269

1270

1271

1272

1273

References

1. J. Beatty, Weighing the Risks: Stalemate in the Classical/Balance Controversy. *Journal of the History of Biology* **20**, 289–319 (1987).

2. A. D. Kern, M. W. Hahn, The Neutral Theory in Light of Natural Selection. *Molecular Biology and Evolution* **35**, 1366–1371 (2018).

3. J. D. Jensen, B. A. Payseur, W. Stephan, C. F. Aquadro, M. Lynch, D. Charlesworth, B. Charlesworth, The importance of the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution* **73**, 111–114 (2019).

4. H. J. Muller, Our load of mutations. *American Journal of Human Genetics* **2**, 111–176 (1950).

5. M. Kimura, Evolutionary Rate at the Molecular Level. *Nature* **217**, 624–626 (1968).

6. J. L. King, T. H. Jukes, Non-Darwinian Evolution. *Science* **164**, 788–798 (1969).

7. M. Kimura, T. Ohta, On some principles governing molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 2848–52 (1974).

8. M. Kimura, Model of Effectively Neutral Mutations in Which Selective Constraint is Incorporated. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 3440–3444 (1979).

9. M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, 1983).

10. E. V. Koonin, Splendor and misery of adaptation, or the importance of neutral null for understanding evolution. *BMC Biol* **14**, 114 (2016).

11. T. Dobzhansky, A Review of Some Fundamental Concepts and Problems of Population Genetics. *Cold Spring Harb Symp Quant Biol* **20**, 1–15 (1955).

12. A. C. Allison, The distribution of the sickle-cell trait in East Africa and elsewhere, and its apparent relationship to the incidence of subtertian malaria. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **48**, 312–318 (1954).

13. P. A. Parsons, W. F. Bodmer, The Evolution of Overdominance: Natural Selection and Heterozygote Advantage. *Nature* **190**, 7–12 (1961).

14. R. C. Lewontin, A General Method for Investigating the Equilibrium of Gene Frequency in a Population. *Genetics* **43**, 419–434 (1958).

15. G. C. Williams, *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton University Press, Princeton, N.J, Facsim. ed., with new pref., 1996)*Princeton science library*.

1307 16. G. Bell, Fluctuating selection: the perpetual renewal of adaptation in variable
1308 environments. *Philos Trans R Soc Lond B Biol Sci* **365**, 87–97 (2010).

1309 17. B. Charlesworth, D. Charlesworth, Neutral Variation in the Context of Selection. *Mol Biol*
1310 *Evol* **35**, 1359–1361 (2018).

1311 18. H. Akashi, S. W. Schaeffer, Natural Selection and the Frequency Distributions of "Silent"
1312 DNA Polymorphism in Drosophila. *Genetics* **146**, 295–307 (1997).

1313 19. I. M. Ehrenreich, J. Bloom, N. Torabi, X. Wang, Y. Jia, L. Kruglyak, Genetic Architecture of
1314 Highly Complex Chemical Resistance Traits across Four Yeast Strains. *PLoS Genet* **8**,
1315 e1002570 (2012).

1316 20. J. Peter, M. De Chiara, A. Friedrich, J. Yue, D. Pflieger, A. Bergström, A. Sigwalt, B. Barre,
1317 K. Freel, A. Llored, C. Cruaud, K. Labadie, J.-M. Aury, B. Istace, K. Lebrigand, P. Barbry,
1318 S. Engelen, A. Lemainque, P. Wincker, G. Liti, J. Schacherer, Genome evolution across
1319 1,011 Saccharomyces cerevisiae isolates. *Nature* **556**, 339–344 (2018).

1320 21. J. S. Bloom, J. Boocock, S. Treusch, M. J. Sadhu, L. Day, H. Oates-Barker, L. Kruglyak,
1321 Rare variants contribute disproportionately to quantitative trait variation in yeast. *eLife* **8**,
1322 e49212 (2019).

1323 22. S. Rahman, S. L. Kosakovsky Pond, A. Webb, J. Hey, Weak selection on synonymous
1324 codons substantially inflates dN/dS estimates in bacteria. *Proceedings of the National*
1325 *Academy of Sciences* **118**, e2023575118 (2021).

1326 23. D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A. G. Taliun,
1327 A. Corvelo, S. M. Gogarten, H. M. Kang, A. N. Pitsillides, J. LeFaive, S. Lee, X. Tian, B. L.
1328 Browning, S. Das, A.-K. Emde, W. E. Clarke, D. P. Loesch, A. C. Shetty, T. W. Blackwell,
1329 A. V. Smith, Q. Wong, X. Liu, M. P. Conomos, D. M. Bobo, F. Aguet, C. Albert, A. Alonso,
1330 K. G. Ardlie, D. E. Arking, S. Aslibekyan, P. L. Auer, J. Barnard, R. G. Barr, L. Barwick, L.
1331 C. Becker, R. L. Beer, E. J. Benjamin, L. F. Bielak, J. Blangero, M. Boehnke, D. W.
1332 Bowden, J. A. Brody, E. G. Burchard, B. E. Cade, J. F. Casella, B. Chalazan, D. I.
1333 Chasman, Y.-D. I. Chen, M. H. Cho, S. H. Choi, M. K. Chung, C. B. Clish, A. Correa, J. E.
1334 Curran, B. Custer, D. Darbar, M. Daya, M. de Andrade, D. L. DeMeo, S. K. Dutcher, P. T.
1335 Ellinor, L. S. Emery, C. Eng, D. Fatkin, T. Fingerlin, L. Forer, M. Fornage, N. Franceschini,
1336 C. Fuchsberger, S. M. Fullerton, S. Germer, M. T. Gladwin, D. J. Gottlieb, X. Guo, M. E.
1337 Hall, J. He, N. L. Heard-Costa, S. R. Heckbert, M. R. Irvin, J. M. Johnsen, A. D. Johnson,
1338 R. Kaplan, S. L. R. Kardia, T. Kelly, S. Kelly, E. E. Kenny, D. P. Kiel, R. Klemmer, B. A.
1339 Konkle, C. Kooperberg, A. Köttgen, L. A. Lange, J. Lasky-Su, D. Levy, X. Lin, K.-H. Lin, C.
1340 Liu, R. J. F. Loos, L. Garman, R. Gerszten, S. A. Lubitz, K. L. Lunetta, A. C. Y. Mak, A.
1341 Manichaikul, A. K. Manning, R. A. Mathias, D. D. McManus, S. T. McGarvey, J. B. Meigs,
1342 D. A. Meyers, J. L. Mikulla, M. A. Minear, B. D. Mitchell, S. Mohanty, M. E. Montasser, C.
1343 Montgomery, A. C. Morrison, J. M. Murabito, A. Natale, P. Natarajan, S. C. Nelson, K. E.
1344 North, J. R. O'Connell, N. D. Palmer, N. Pankratz, G. M. Peloso, P. A. Peyser, J. Pleiness,
1345 W. S. Post, B. M. Psaty, D. C. Rao, S. Redline, A. P. Reiner, D. Roden, J. I. Rotter, I.
1346 Ruczinski, C. Sarnowski, S. Schoenherr, D. A. Schwartz, J.-S. Seo, S. Seshadri, V. A.
1347 Sheehan, W. H. Sheu, M. B. Shoemaker, N. L. Smith, J. A. Smith, N. Sotoodehnia, A. M.
1348 Stilp, W. Tang, K. D. Taylor, M. Telen, T. A. Thornton, R. P. Tracy, D. J. Van Den Berg, R.
1349 S. Vasan, K. A. Viaud-Martinez, S. Vrieze, D. E. Weeks, B. S. Weir, S. T. Weiss, L.-C.

1350 Weng, C. J. Willer, Y. Zhang, X. Zhao, D. K. Arnett, A. E. Ashley-Koch, K. C. Barnes, E.
1351 Boerwinkle, S. Gabriel, R. Gibbs, K. M. Rice, S. S. Rich, E. K. Silverman, P. Qasba, W.
1352 Gan, G. J. Papanicolaou, D. A. Nickerson, S. R. Browning, M. C. Zody, S. Zöllner, J. G.
1353 Wilson, L. A. Cupples, C. C. Laurie, C. E. Jaquish, R. D. Hernandez, T. D. O'Connor, G. R.
1354 Abecasis, Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.
1355 *Nature* **590**, 290–299 (2021).

1356 24. R. S. Dhindsa, Q. Wang, D. Vitsios, O. S. Burren, F. Hu, J. E. DiCarlo, L. Kruglyak, D. G.
1357 MacArthur, M. E. Hurles, S. Petrovski, A minimal role for synonymous variation in human
1358 disease. *The American Journal of Human Genetics* **109**, 2105–2109 (2022).

1359 25. X. Guo, A. Chavez, A. Tung, Y. Chan, C. Kaas, Y. Yin, R. Cecchi, S. L. Garnier, E. D.
1360 Kelsic, M. Schubert, J. E. DiCarlo, J. J. Collins, G. M. Church, High-throughput creation
1361 and functional profiling of DNA sequence variant libraries using CRISPR–Cas9 in yeast.
1362 *Nature Biotechnology* **36**, 195776 (2018).

1363 26. E. Sharon, S. A. Chen, N. M. Khosla, J. D. Smith, J. K. Pritchard, H. B. Fraser, Functional
1364 Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell* **175**, 544-
1365 557.e16 (2018).

1366 27. K. R. Roy, J. D. Smith, S. C. Vonesch, G. Lin, C. S. Tu, A. R. Lederer, A. Chu, S. Suresh,
1367 M. Nguyen, J. Horecka, A. Tripathi, W. T. Burnett, M. A. Morgan, J. Schulz, K. M. Orsley,
1368 W. Wei, R. S. Aiyar, R. W. Davis, V. A. Bankaitis, J. E. Haber, M. L. Salit, R. P. St.Onge, L.
1369 M. Steinmetz, Multiplexed precision genome editing with trackable genomic barcodes in
1370 yeast. *Nature Biotechnology* **36** (2018).

1371 28. M. J. Sadhu, J. S. Bloom, L. Day, J. J. Siegel, S. Kosuri, L. Kruglyak, I. Müller, Highly
1372 parallel genome variant engineering with CRISPR–Cas9. *Nature Genetics* **50**, 510–514
1373 (2018).

1374 29. S.-A. A. Chen, A. F. Kern, R. M. L. Ang, Y. Xie, H. B. Fraser, Gene-by-environment
1375 interactions are pervasive among natural genetic variants. *Cell Genomics* **3**, 100273
1376 (2023).

1377 30. R. M. L. Ang, S.-A. A. Chen, A. F. Kern, Y. Xie, H. B. Fraser, Widespread epistasis among
1378 beneficial genetic variants revealed by high-throughput genome editing. *Cell Genomics* **3**,
1379 100260 (2023).

1380 31. M. Conrad, J. Schothorst, H. N. Kankipati, G. Van Zeebroeck, M. Rubio-Texeira, J. M.
1381 Thevelein, Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae. *FEMS*
1382 *Microbiology Reviews* **38**, 254–299 (2014).

1383 32. J. R. Broach, Nutritional control of growth and development in yeast. *Genetics* **192**, 73–105
1384 (2012).

1385 33. R. Loewith, M. N. Hall, Target of rapamycin (TOR) in nutrient signaling and growth control.
1386 *Genetics* **189**, 1177–201 (2011).

1387 34. D. A. Hogan, P. Sundstrom, The Ras/cAMP/PKA signaling pathway and virulence in
1388 Candida albicans. *Future Microbiology* **4**, 1263–1270 (2009).

1389 35. D. R. Pentland, E. Piper-Brown, F. A. Mühlschlegel, C. W. Gourlay, Ras signalling in
1390 pathogenic yeasts. *Microbial Cell* **5**, 63–73 (2017).

1391 36. R. Tisi, F. Belotti, E. Martegani, "Yeast as a Model for Ras Signalling" in *Ras Signaling*, L.
1392 Trabalzini, S. F. Retta, Eds. (Humana Press, Totowa, NJ, 2014;
1393 https://link.springer.com/10.1007/978-1-62703-791-4_23)vol. 1120 of *Methods in Molecular*
1394 *Biology*, pp. 359–390.

1395 37. K. C. Kao, G. Sherlock, Molecular characterization of clonal interference during adaptive
1396 evolution in asexual populations of Saccharomyces cerevisiae. *Nature genetics* **40**, 1499–
1397 504 (2008).

1398 38. J. W. Wenger, J. Piotrowski, S. Nagarajan, K. Chiotti, G. Sherlock, F. Rosenzweig, Hunger
1399 artists: Yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLoS*
1400 *Genetics* **7** (2011).

1401 39. D. J. Kvitek, G. Sherlock, Whole genome, whole population sequencing reveals that loss of
1402 signaling networks is the major adaptive strategy in a constant environment. *PLoS*
1403 *genetics* **9**, e1003972 (2013).

1404 40. J. Hong, D. Gresham, Molecular Specificity, Convergence and Constraint Shape Adaptive
1405 Evolution in Nutrient-Poor Environments. *PLoS Genetics* **10** (2014).

1406 41. T. K. Sato, M. Tremaine, L. S. Parreiras, A. S. Hebert, K. S. Myers, A. J. Higbee, M. Sardi,
1407 S. J. McIlwain, I. M. Ong, R. J. Breuer, R. Avanasi Narasimhan, M. A. McGee, Q.
1408 Dickinson, A. La Reau, D. Xie, M. Tian, J. L. Reed, Y. Zhang, J. J. Coon, C. T. Hittinger, A.
1409 P. Gasch, R. Landick, Directed Evolution Reveals Unexpected Epistatic Interactions That
1410 Alter Metabolic Regulation and Enable Anaerobic Xylose Use by Saccharomyces
1411 cerevisiae. *PLoS Genetics* **12**, 1–31 (2016).

1412 42. S. Venkataram, B. Dunn, Y. Li, A. Agarwala, J. Chang, E. R. Ebel, K. Geiler-Samerotte, L.
1413 Hérissant, J. R. Blundell, S. F. Levy, D. S. Fisher, G. Sherlock, D. A. Petrov, Development
1414 of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast.
1415 *Cell* **166**, 1585–1596 (2016).

1416 43. M. B. Taylor, J. Phan, J. T. Lee, M. McCadden, I. M. Ehrenreich, Diverse genetic
1417 architectures lead to the same cryptic phenotype in a yeast cross. *Nature Communications*
1418 **7**, 1–6 (2016).

1419 44. J. A. Granek, D. Murray, Ö. Kayrkçi, P. M. Magwene, The genetic architecture of biofilm
1420 formation in a clinical isolate of Saccharomyces cerevisiae. *Genetics* **193**, 587–600 (2013).

1421 45. A. Yadav, A. Radhakrishnan, G. Bhanot, H. Sinha, Differential regulation of antagonistic
1422 pleiotropy in synthetic and natural populations suggests its role in adaptation. *G3* **5**, 699–
1423 709 (2015).

1424 46. Y. Li, S. Venkataram, A. Agarwala, B. Dunn, D. A. Petrov, G. Sherlock, D. S. Fisher,
1425 Hidden Complexity of Yeast Adaptation under Simple Evolutionary Conditions. *Current*
1426 *biology : CB* **28**, 515-525.e6 (2018).

1427 47. C. M. Jakobson, J. Hartl, P. Trébulle, M. Mülleder, D. F. Jarosz, M. Ralser, A genome-to-
1428 proteome atlas charts natural variants controlling proteome diversity and forecasts their
1429 fitness effects. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.10.18.619054.

1430 48. S. Wright, "The roles of mutation, inbreeding, crossbreeding and selection in evolution" in
1431 *Proceedings of the Sixth International Congress on Genetics*, D. F. Jones, Ed. (The
1432 Genetics Society of America, 1932)vol. 1, pp. 356–366.

1433 49. J.-L. Legras, V. Galeote, F. Bigey, C. Camarasa, S. Marsit, T. Nidelet, I. Sanchez, A.
1434 Couloux, J. Guy, R. Franco-Duarte, M. Marcet-Houben, T. Gabaldon, D. Schuller, J. P.
1435 Sampaio, S. Dequin, Adaptation of S. cerevisiae to Fermented Food Environments
1436 Reveals Remarkable Genome Plasticity and the Footprints of Domestication. *Mol Biol Evol*
1437 **35**, 1712–1727 (2018).

1438 50. M. C. Bassalo, A. D. Garst, A. Choudhury, W. C. Grau, E. J. Oh, E. Spindler, T. Lipscomb,
1439 R. T. Gill, Deep scanning lysine metabolism in Escherichia coli. *Molecular systems biology*
1440 **14**, e8371 (2018).

1441 51. C. Verduyn, E. Postma, W. A. Scheffers, J. P. Van Dijken, Effect of benzoic acid on
1442 metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and
1443 alcoholic fermentation. *Yeast* **8**, 501–517 (1992).

1444 52. M. Razo-Mejia, M. Mani, D. Petrov, Bayesian inference of relative fitness on high-
1445 throughput pooled competition assays. *PLOS Computational Biology* **20**, e1011937 (2024).

1446 53. S. Li, A. Balmain, C. M. Counter, A model for RAS mutation patterns in cancers: finding the
1447 sweet spot. *Nature Reviews Cancer*, 1 (2018).

1448 54. T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I.
1449 Furman, Y. Pilpel, An Evolutionarily Conserved Mechanism for Controlling the Efficiency of
1450 Protein Translation. *Cell* **141**, 344–354 (2010).

1451 55. P. C. Ng, S. Henikoff, Predicting deleterious amino acid substitutions. *Genome Res* **11**,
1452 863–874 (2001).

1453 56. P. C. Ng, S. Henikoff, Accounting for Human Polymorphisms Predicted to Affect Protein
1454 Function. *Genome Res.* **12**, 436–446 (2002).

1455 57. J. Echave, S. J. Spielman, C. O. Wilke, Causes of evolutionary rate variation among
1456 protein sites. *Nat Rev Genet* **17**, 109–121 (2016).

1457 58. J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, The FoldX web
1458 server: an online force field. *Nucleic Acids Research* **33**, W382–W388 (2005).

1459 59. Y. O. Zhu, G. Sherlock, D. A. Petrov, Extremely Rare Polymorphisms in Saccharomyces
1460 cerevisiae Allow Inference of the Mutational Spectrum. *PLoS Genetics* **13**, 1–16 (2017).

1461 60. B. Gallone, J. Steensels, T. Prahl, L. Soriaga, V. Saels, B. Herrera-Malaver, A. Merlevede,
1462 M. Roncoroni, K. Voordeckers, L. Miraglia, C. Teiling, B. Steffy, M. Taylor, A. Schwartz, T.
1463 Richardson, C. White, G. Baele, S. Maere, K. J. Verstrepen, Domestication and
1464 Divergence of Saccharomyces cerevisiae Beer Yeasts. *Cell* **166**, 1397-1410.e16 (2016).

61. I. Belda, S. Izquierdo-Gea, B. Benitez-Dominguez, J. Ruiz, J. C. C. Vila, Wine Fermentation as a Model System for Microbial Ecology and Evolution. *Environmental Microbiology* **27**, e70092 (2025).

62. M. De Chiara, B. P. Barré, K. Persson, A. Irizar, C. Vischioni, S. Khaiwal, S. Stenberg, O. C. Amadi, G. Žun, K. Doberšek, C. Taccioli, J. Schacherer, U. Petrovič, J. Warringer, G. Liti, Domestication reprogrammed the budding yeast life cycle. *Nat Ecol Evol* **6**, 448–460 (2022).

63. R. Tengölics, B. Szappanos, M. Mülleder, D. Kalapis, G. Grézal, C. Sajben, F. Agostini, J. B. Mokochinski, B. Bálint, L. G. Nagy, M. Ralser, B. Papp, The metabolic domestication syndrome of budding yeast. *Proceedings of the National Academy of Sciences* **121**, e2313354121 (2024).

64. N. Condic, H. Amiji, D. Patel, W. C. Shropshire, N. O. Lermi, Y. Sabha, B. John, B. Hanson, G. I. Karras, Selection for robust metabolism in domesticated yeasts is driven by adaptation to Hsp90 stress. *Science*, doi: 10.1126/science.adi3048 (2024).

65. M. Plech, J. A. G. M. de Visser, R. Korona, Heterosis Is Prevalent Among Domesticated but not Wild Strains of Saccharomyces cerevisiae. *G3 Genes|Genomes|Genetics* **4**, 315–323 (2014).

66. S.-F. Duan, J.-Y. Shi, Q. Yin, R.-P. Zhang, P.-J. Han, Q.-M. Wang, F.-Y. Bai, Reverse Evolution of a Classic Gene Network in Yeast Offers a Competitive Advantage. *Current Biology* **29**, 1126-1136.e5 (2019).

67. I. Cvijović, B. H. Good, M. M. Desai, The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics* **209**, 1235–1278 (2018).

68. N. Dukler, M. R. Mughal, R. Ramani, Y.-F. Huang, A. Siepel, Extreme purifying selection against point mutations in the human genome. *Nat Commun* **13**, 4312 (2022).

69. S. Richards, N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H. L. Rehm, Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405–423 (2015).

70. N. Whiffin, E. Minikel, R. Walsh, A. H. O'Donnell-Luria, K. Karczewski, A. Y. Ing, P. J. R. Barton, B. Funke, S. A. Cook, D. MacArthur, J. S. Ware, Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med* **19**, 1151–1158 (2017).

71. M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M.

Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

72. S. Gudmundsson, M. Singer-Berk, N. A. Watts, W. Phu, J. K. Goodrich, M. Solomonson, G. A. D. Consortium, H. L. Rehm, D. G. MacArthur, A. O'Donnell-Luria, Variant interpretation using population databases: Lessons from gnomAD. *Human Mutation* **43**, 1012–1030 (2022).

73. M. Plank, Interaction of TOR and PKA Signaling in S. cerevisiae. *Biomolecules* **12**, 210 (2022).

74. J. Pinto, C. Haberkorn, M. Franzén, A. J. M. Tack, R. Stelkens, Fermentative Yeast Diversity at the Northern Range Limit of Their Oak Tree Hosts. *Environmental Microbiology Reports* **17**, e70110 (2025).

75. L. Riles, J. C. Fay, Genetic Basis of Variation in Heat and Ethanol Tolerance in Saccharomyces cerevisiae. *G3 Genes|Genomes|Genetics* **9**, 179–188 (2019).

76. S.-F. Duan, P.-J. Han, Q.-M. Wang, W.-Q. Liu, J.-Y. Shi, K. Li, X.-L. Zhang, F.-Y. Bai, The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nat Commun* **9**, 2690 (2018).

77. J. Boocock, M. J. Sadhu, A. Durvasula, J. S. Bloom, L. Kruglyak, Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast. *Science* **371**, 415–419 (2021).

78. D. Aggeli, D. A. Marad, X. Liu, S. W. Buskirk, S. F. Levy, G. I. Lang, Overdominant and partially dominant mutations drive clonal adaptation in diploid Saccharomyces cerevisiae. *Genetics* **221**, iyac061 (2022).

79. A. B. Sunshine, C. Payen, G. T. Ong, I. Liachko, K. M. Tan, M. J. Dunham, The Fitness Consequences of Aneuploidy Are Driven by Condition-Dependent Gene Effects. *PLOS Biology* **13**, e1002155 (2015).

80. S. M. Ardell, A. Martsul, M. S. Johnson, S. Kryazhimskiy, Environment-independent distribution of mutational effects emerges from microscopic epistasis. *Science* **386**, 87–92 (2024).

81. C. Buzby, Y. Plavskin, F. M. O. Sartori, Q. Tong, J. K. Vail, M. L. Siegal, Epistasis and cryptic QTL identified using modified bulk segregant analysis of copper resistance in budding yeast. *Genetics* **229**, iyaf026 (2025).

82. KhaiwalSakshi, D. ChiaraMatteo, B. P, Barrio-HernandezInigo, StenbergSimon, BeltraoPedro, WarringerJonas, LitiGianni, Predicting natural variation in the yeast phenotypic landscape with machine learning. *Molecular Systems Biology*, doi: 10.1038/s44320-025-00136-y (2025).

1543 83. S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, G. Sherlock,
1544 Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**,
1545 181–6 (2015).

1546 84. Y. Li, D. A. Petrov, G. Sherlock, Single nucleotide mapping of trait space reveals Pareto
1547 fronts that constrain adaptation. *Nature Ecology and Evolution* **3**, 1539–1551 (2019).

1548 85. G. Kinsler, K. Geiler-Samerotte, D. Petrov, Fitness variation across subtle environmental
1549 perturbations reveals local modularity and global pleiotropy of adaptation. *eLife* **9**, 1–52
1550 (2020).

1551 86. D. Aggeli, Y. Li, G. Sherlock, Changes in the distribution of fitness effects and adaptive
1552 mutational spectra following a single first step towards adaptation. *Nat Commun* **12**, 5193
1553 (2021).

1554 87. M. Hays, K. Schwartz, D. T. Schmidtke, D. Aggeli, G. Sherlock, Paths to adaptation under
1555 fluctuating nitrogen starvation: The spectrum of adaptive mutations in Saccharomyces
1556 cerevisiae is shaped by retrotransposons and microhomology-mediated recombination.
1557 *PLoS Genet* **19**, e1010747 (2023).

1558 88. K. Schmidlin, S. Apodaca, D. Newell, A. Sastokas, G. Kinsler, K. Geiler-Samerotte,
1559 Distinguishing mutants that resist drugs via different mechanisms by examining fitness
1560 tradeoffs. *eLife* **13**, RP94144 (2024).

1561 89. L. Benatuil, J. M. Perez, J. Belk, C.-M. Hsieh, An improved yeast transformation method
1562 for the generation of very large human antibody libraries. *Protein Engineering, Design and*
1563 *Selection* **23**, 155–159 (2010).

1564 90. J. D. Boeke, J. Trueheart, G. Natsoulis, G. R. Fink, "[10] 5-Fluoroorotic acid as a selective
1565 agent in yeast molecular genetics" in *Methods in Enzymology* (Elsevier, 1987;
1566 https://linkinghub.elsevier.com/retrieve/pii/0076687987540769)vol. 154, pp. 164–175.

1567 91. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads.
1568 *EMBnet.journal* **17**, 10–12 (2011).

1569 92. T. Magoč, S. L. Salzberg, FLASH: fast length adjustment of short reads to improve
1570 genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

1571 93. W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. S. Ritchie, A. Thormann, P. Flicek, F.
1572 Cunningham, The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).

1573 94. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K.
1574 Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl,
1575 A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S.
1576 Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T.
1577 Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli,
1578 D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–
1579 589 (2021).

1580 95. M. Varadi, D. Bertoni, P. Magana, U. Paramval, I. Pidruchna, M. Radhakrishnan, M.
1581 Tsenkov, S. Nair, M. Mirdita, J. Yeo, O. Kovalevskiy, K. Tunyasuvunakool, A. Laydon, A.

Žídek, H. Tomlinson, D. Hariharan, J. Abrahamson, T. Green, J. Jumper, E. Birney, M. Steinegger, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research* **52**, D368–D375 (2024).

96. W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

97. R. P. Joosten, T. A. H. Te Beek, E. Krieger, M. L. Hekkelman, R. W. W. Hooft, R. Schneider, C. Sander, G. Vriend, A series of PDB related databases for everyday needs. *Nucleic Acids Research* **39**, D411–D419 (2011).

98. P. Beltrao, D. Memon, M. Galardini, B. P. Busby, A. Typas, O. Wagih, A resource of variant effect predictions of single nucleotide variants in model organisms. *Molecular Systems Biology* **14**, e8430 (2018).

99. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).

1613 **Supplementary Table S1:** Genes considered in this study

Standard name	Systematic name	Pathway
BCY1	YIL033C	Ras/PKA
CDC25	YLR310C	Ras/PKA
CYR1	YJL005W	Ras/PKA
GPA2	YER020W	Ras/PKA
GPB1	YOR371C	Ras/PKA
GPB2	YAL056W	Ras/PKA
GPR1	YDL035C	Ras/PKA
IRA1	YBR140C	Ras/PKA
IRA2	YOL081W	Ras/PKA
PDE1	YGL248W	Ras/PKA
PDE2	YOR360C	Ras/PKA
RAS1	YOR101W	Ras/PKA
RAS2	YNL098C	Ras/PKA
RGS2	YOR107W	Ras/PKA
TFS1	YLR178C	Ras/PKA
TPK1	YJL164C	Ras/PKA
TPK2	YPL203W	Ras/PKA
TPK3	YKL166C	Ras/PKA
KOG1	YHR186C	TOR/Sch9
LST8	YNL006W	TOR/Sch9
SCH9	YHR205W	TOR/Sch9
TCO89	YPL180W	TOR/Sch9
TOR1	YJR066W	TOR/Sch9
TOR2	YKL203C	TOR/Sch9

1614

1615