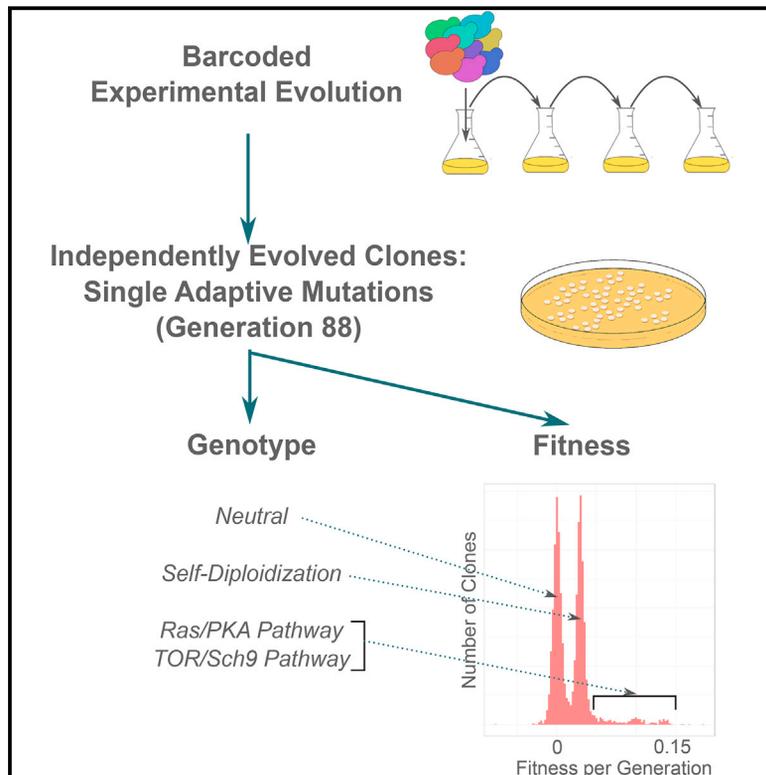


# Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast

## Graphical Abstract



## Authors

Sandeep Venkataram, Barbara Dunn, Yuping Li, ..., Daniel S. Fisher, Gavin Sherlock, Dmitri A. Petrov

## Correspondence

gsherloc@stanford.edu (G.S.), dpetrov@stanford.edu (D.A.P.)

## In Brief

How can the entire suite of adaptive mutations that occur in an evolving population be pinpointed and each of their fitness effects measured?

## Highlights

- High-throughput method to survey adaptive mutations in laboratory evolution
- Isolation of thousands of independent single adaptive events genome-wide
- Bulk fitness measurements and genotyping generates a genotype-to-fitness map
- Large sample size allows sensitive detection of fitness differences even within genes

# Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast

Sandeep Venkataram,<sup>1,7</sup> Barbara Dunn,<sup>2,7</sup> Yuping Li,<sup>1</sup> Atish Agarwala,<sup>3</sup> Jessica Chang,<sup>2</sup> Emily R. Ebel,<sup>1</sup> Kerry Geiler-Samerotte,<sup>1</sup> Lucas Hérisant,<sup>2</sup> Jamie R. Blundell,<sup>4,5</sup> Sasha F. Levy,<sup>5,6</sup> Daniel S. Fisher,<sup>1,4</sup> Gavin Sherlock,<sup>2,\*</sup> and Dmitri A. Petrov<sup>1,8,\*</sup>

<sup>1</sup>Department of Biology

<sup>2</sup>Department of Genetics

<sup>3</sup>Department of Physics

<sup>4</sup>Department of Applied Physics

Stanford University, Stanford, CA 94305, USA

<sup>5</sup>Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794-5252, USA

<sup>6</sup>Department of Biochemistry and Cellular Biology, Stony Brook University, Stony Brook, NY 11794-5215, USA

<sup>7</sup>Co-first author

<sup>8</sup>Lead Contact

\*Correspondence: [gsherloc@stanford.edu](mailto:gsherloc@stanford.edu) (G.S.), [dpetrov@stanford.edu](mailto:dpetrov@stanford.edu) (D.A.P.)

<http://dx.doi.org/10.1016/j.cell.2016.08.002>

## SUMMARY

Adaptive evolution plays a large role in generating the phenotypic diversity observed in nature, yet current methods are impractical for characterizing the molecular basis and fitness effects of large numbers of individual adaptive mutations. Here, we used a DNA barcoding approach to generate the genotype-to-fitness map for adaptation-driving mutations from a *Saccharomyces cerevisiae* population experimentally evolved by serial transfer under limiting glucose. We isolated and measured the fitness of thousands of independent adaptive clones and sequenced the genomes of hundreds of clones. We found only two major classes of adaptive mutations: self-diploidization and mutations in the nutrient-responsive Ras/PKA and TOR/Sch9 pathways. Our large sample size and precision of measurement allowed us to determine that there are significant differences in fitness between mutations in different genes, between different paralogs, and even between different classes of mutations within the same gene.

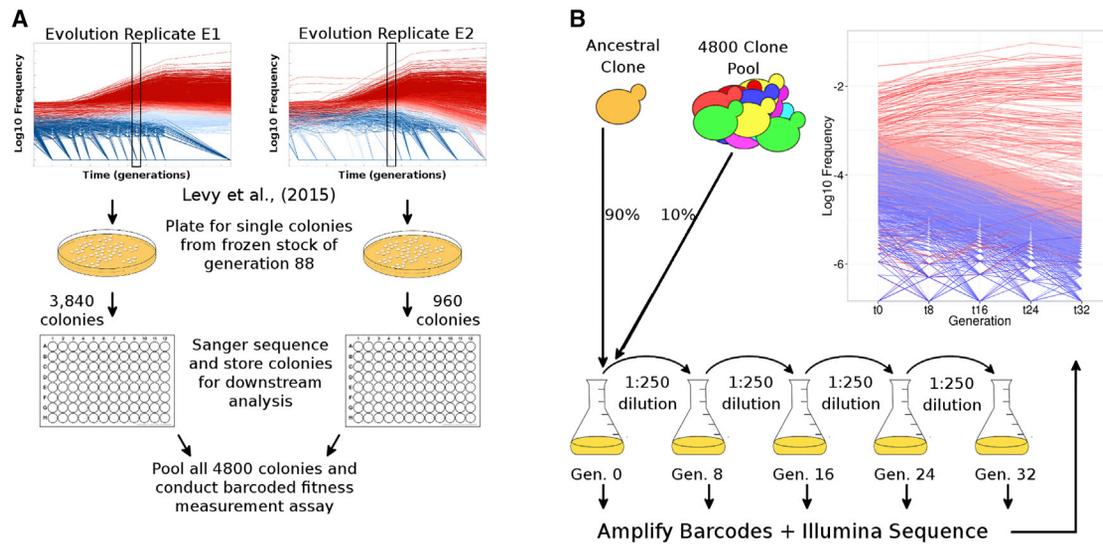
## INTRODUCTION

Adaptive evolution is a major driving force behind the observed phenotypic diversity in nature (Darwin, 1872; reviewed in Givnish, 2015; Soulebeau et al., 2015) and is of key importance to many problems of biomedical interest, including cancer (Greaves and Maley, 2012; Korolev et al., 2014; Landau et al., 2013; Nowell, 1976) and the emergence of drug resistance (Davies and Davies, 2010; Palmer and Kishony, 2013; Pennings, 2012; Toprak et al., 2011). To further understand the process

of adaptation, it is essential to obtain a large, statistically representative number of individual adaptive events and determine their fitness effects and molecular nature.

While there are many methods for identifying instances of adaptive evolution in natural populations, they are not suitable for a comprehensive analysis of the spectrum of mutations that drive adaptation. Indeed, methods that infer selection in natural populations (reviewed in Lachance and Tishkoff, 2013; Oleksyk et al., 2010; Stinchcombe and Hoekstra, 2008; Vitti et al., 2013) are typically unable to identify adaptive mutations with single base-pair resolution, much less quantify the fitness effects of single adaptive mutations. Mechanistic studies can be conducted in genetically tractable systems where one can measure the fitness effects of a set of engineered mutations (Bank et al., 2015; Bozek et al., 2014; Fowler and Fields, 2014; Giaever et al., 2002; Hietpas et al., 2013; De Meester et al., 2002; Rich et al., 2016; Sliwa and Korona, 2005; Warringer et al., 2011; Weinreich et al., 2006). However, mutations studied in such systems are typically limited to a small, artificial, and predominantly deleterious subset of possible mutations, e.g., whole-gene knockout mutations or deep mutational scanning of one or a few genomic regions.

In principle, microbial experimental evolution provides an excellent framework for the comprehensive study of adaptive mutations due to the ease of both identifying adaptive mutations and assaying their fitness by pairwise competition. Two experimental evolution approaches for identifying large numbers of independent beneficial mutations are to either sequence multiple isolates from populations evolved under identical conditions (e.g., Barrick et al., 2009; Gresham et al., 2008; Kryazhimskiy et al., 2014; Kvittek and Sherlock, 2011; Tenaillon et al., 2012; reviewed in Dettman et al., 2012; Long et al., 2015), or to conduct whole-population, whole-genome sequencing at multiple time points during the evolution (Herron and Doebeli, 2013; Kvittek and Sherlock, 2013; Lang et al., 2013). However, these approaches are limited to identifying only a subset of



**Figure 1. Experimental Procedures to Select and Measure Fitness of Evolved Clones**

(A) Schematic of isolation and identification of individual evolved yeast clones. We isolated 4,800 single colonies from generation 88 across both replicate evolution experiments from [Levy et al. \(2015\)](#), determined their lineage barcodes, and stored them individually.

(B) Schematic of barcoded fitness measurement assay. We grew all 4,800 colonies individually (not shown) and pooled them. The pool was mixed with an ancestral clone at a 1:9 ratio and the mixture was propagated for 32 generations in four independent batches (two to three replicates per batch). At each transfer (every eight generations), we isolated DNA, amplified the barcodes, and conducted high-throughput sequencing to estimate the frequency trajectory of each barcode. The inset graph shows the frequency trajectory of all lineages with fitness > 1%, where adaptive lineages are colored in red (darker red lineages are more fit) and neutral lineages are colored in blue. Fitness was estimated using 24 generations of data from these frequency trajectories (see [STAR Methods](#)). Raw data for the sampled clones and their fitness measurements are in [Tables S1, S2, and S3](#).

See also [Figure S6](#) and [Data S1](#).

high-frequency and easy to sequence mutations. Moreover, separating the adaptive mutations from those that are merely hitchhiking remains a challenge ([Voordeckers and Verstrepen, 2015](#)). For example, in many studies the sequenced clones were isolated after hundreds or thousands of generations to ensure the presence of adaptive mutations, resulting in multiple mutations per clone ([Barrick et al., 2009](#); [Kryazhimskiy et al., 2014](#); [Tenailon et al., 2012](#)). This makes it difficult to distinguish adaptive mutations from hitchhikers and also precludes the measurement of the fitness effects of individual beneficial mutations in isolation. By contrast, whole-population genome sequencing provides us only with the trajectories of easy to sequence mutations that rise to high frequencies (>1%), at which time they tend to be present in clones with multiple mutations, and their behavior is driven by complex clonal interference dynamics ([Desai and Fisher, 2007](#); [Herron and Doebeli, 2013](#); [Kvitek and Sherlock, 2013](#)); this prevents both the identification of very low-frequency yet beneficial mutations and the precise estimation of their individual or marginal selective effects. Finally, fitness measurements are typically done in a low throughput, pairwise fashion, precluding generation of a comprehensive genotype-to-fitness map.

Here, we use our lineage tracking method ([Levy et al., 2015](#)) to solve these technological limitations and characterize both the genetic basis and fitness effects of hundreds of independent adaptive mutations in a laboratory evolution experiment using *Saccharomyces cerevisiae*. Using DNA barcodes as neutral markers to track the frequencies of ~500,000 independent line-

ages during an evolution experiment, [Levy et al. \(2015\)](#) identified ~25,000 lineages that gained an adaptive mutation within the first 168 generations of evolution. We have now isolated thousands of clones from a single early time point in those experiments—a point at which we expect most adaptive lineages to carry single adaptive mutations—and identified their DNA barcodes. We then pooled these clones and monitored their barcode frequencies during short-term pooled growth. This allowed us to assign a fitness value to each of the clones, within the context of a single experiment. We then selected and sequenced the genomes of hundreds of known adaptive clones with varying fitness effects, as well as many neutral clones. Combining the sequencing and fitness measurements, we linked the molecular targets of adaptation to their fitness effects and thus built a comprehensive genotype-to-fitness map of the mutations that drove initial adaptive evolution in this system. Our results show that initial adaptation under these conditions is overwhelmingly driven by two distinct classes of mutations, which together explain the bimodal distribution of fitness effects observed in [Levy et al. \(2015\)](#).

## RESULTS

### Isolation of Thousands of Evolved Clones and Parallel Measurement of Their Fitness

We isolated 4,800 random, single-colony-derived clones from frozen population samples taken at generation 88 from the [Levy et al. \(2015\)](#) experimental evolutions ([Figure 1A](#); [Table S1](#)): 3,840 clones were from evolution replicate E1 and 960 clones

from replicate E2. Those evolutions were performed by serial transfer in limiting glucose conditions, such that the populations grew for eight generations each 48 hr growth/dilution cycle. The sampling generation and number of clones were chosen specifically to both maximize the fraction of clones with only a single adaptive mutation and allow fitness measurement assays to be cost-effective (see [STAR Methods](#)). We unambiguously determined the barcode sequence for 4,149 of the clones via Sanger sequencing (see [STAR Methods](#)) and identified 4,009 unique barcodes with 140 duplicates, consistent with random sampling from the [Levy et al. \(2015\)](#) data.

To measure the fitness,  $s$ , of each of these clones, we conducted fitness measurements in a single pooled assay ([Figure 1B](#); [Tables S2](#) and [S3](#)). We grew each of the 4,800 clones independently in liquid media and then pooled equal volumes of their saturated cultures; this pool was then frozen as a stock culture to use for all subsequent fitness measurements described in this work, unless specified otherwise. For each assay, we re-grew the pool from a frozen stock of  $\sim 10^8$  cells, then mixed it 1:9 with a population of the ancestral clone. We then propagated this mixed population by serial transfer through four eight-generation cycles for a total of 32 generations under conditions identical to the original evolution experiment ([Figure 1B](#)); the starting population size for each cycle was  $\sim 5 \times 10^7$  cells, large enough to minimize drift. This design allowed us to measure the fitness relative to the ancestor of each of the 4,800 clones in the pool without allowing substantial further adaptive evolution during the propagation. These measurements were conducted with two to three biological replicates across each of four different experimental batches (experiments conducted on different days).

The frequency of each barcode was measured after each transfer cycle by Illumina sequencing (see [STAR Methods](#)). We detected 3,883 of the 4,009 unique lineage barcodes; clones carrying the 126 missing barcodes may not have recovered from the frozen stock in high enough numbers to establish and thus were not present in the pool used for the fitness measurements. We used the frequency measurements from three of the four eight-generation cycles (for a total of 24 generations of data) to estimate the fitness of the 3,883 clones. Details of the fitness estimation, and extensive analysis of the fitness measurement errors and the batch effects are in the [STAR Methods](#) and [Figures S1](#), [S2](#), [S3](#), [S4](#), and [S5](#). The distribution of fitness effects for all sampled lineages is shown in [Figure S6](#).

The fitness values ( $s$ ) reported throughout this work are the inverse variance-weighted mean and sample SEM across the four batches of fitness measurements and are quoted, following convention, as percent per generation. The fitness measurements are consistent across replicates within batches ([Figures 2A](#) and [S4](#)) and between batches, although not to the same extent ([Figures 2B](#), [2C](#), and [S5](#)). Sources of error between the replicates and batches include counting noise, caused by the growth/bottleneck dynamics of the assay itself, and from sampling and sequencing the DNA from the population, as well as intrinsic experimental noise. In addition, there appear to be systematic deviations among the batches. Batch 2 showed the largest systematic deviations ([Figure S5](#)), on the order of 6.5% for high fitness lineages ( $s > 5\%$ ) rather than the 1%–2% deviations for all other batches ([Figures 2B](#) and [2C](#)), which may be due

to the slightly different measurement protocol used for this batch when compared to the other batches (see [STAR Methods](#)).

Some deviations across the batches might be caused by slight differences in the growth conditions between batches or may be induced by different population compositions during the latter growth cycles of the fitness assay. We considered the possibility that a few lineages present at a substantial frequency in the pool (13 lineages at 1%–8% frequency) could drive non-linear effects at the latter growth cycles of the assay. To investigate this, we created a pool of 500 of the barcoded clones, providing us with a biological replicate of pooling and specifically avoiding the introduction of the anomalously large lineages. We performed the fitness assay as for the larger pool and found that the fitness estimates remained largely unchanged ([Figure 2D](#)) with similar systematic batch deviations, on the order of  $\sim 3.2\%$  for high fitness lineages ( $s > 5\%$ ) (see [STAR Methods](#)). This indicates that most of the among-batch variation is likely to be driven by biological variability and not variation in the pool composition or a few anomalously large lineages. Overall, the systematic effects appear to be small compared to the measured fitness values, and our analysis below controls for the batch effect in all pairwise comparisons.

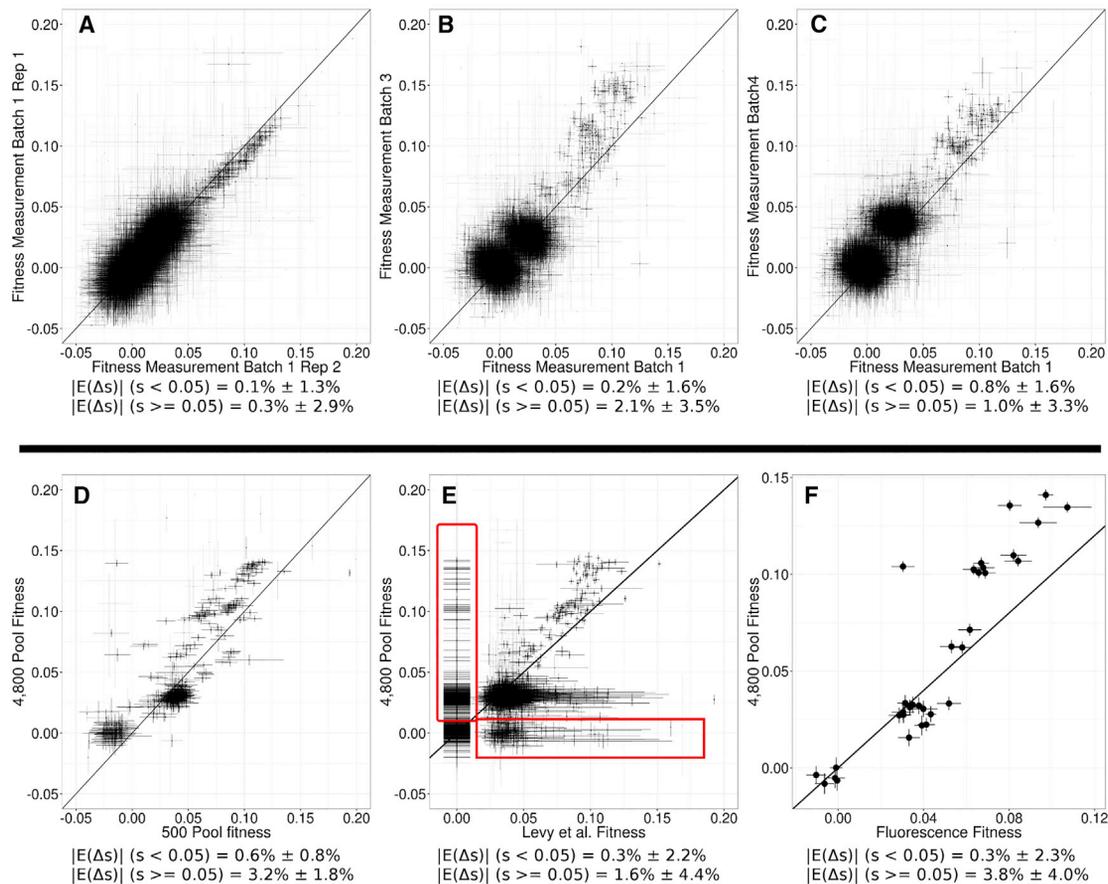
Our fitness measurements are consistent with those of [Levy et al. \(2015\)](#) as reported for both the lineage tracking fitness estimates ([Figure 2E](#)) and pairwise competition assays of single clones against a YFP-marked ancestor ([Figure 2F](#)). We suspect that the deviations in fitness in these assays when compared to our 4,800 pool estimates are largely due to batch effects, although we cannot rule out fitness differences due to frequency-dependent effects as the adaptive clones begin each of these assays at a different starting frequency (see [Levy et al., 2015](#) and [STAR Methods](#) for details).

Note, a number of lineages were classified as adaptive by [Levy et al. \(2015\)](#), while our isolated clones from those lineages proved to be neutral and vice versa (highlighted in [Figure 2E](#)). This is expected: adaptive mutants in lineages called adaptive by [Levy et al. \(2015\)](#) should generally comprise the majority but not all of the cells in their lineages. Thus, there will be instances where the sampled isolate from a lineage does not have the adaptive mutation. Conversely, some sampled isolates from lineages called neutral by [Levy et al. \(2015\)](#) will have acquired an adaptive mutation late enough in the evolution that the lineage was not classified as adaptive. The pooled-clone fitness measurements conducted in this study were thus critical for assigning fitness effects to our isolated clones (see below).

We determined that 59% of our 3,883 sampled lineages were adaptive (defined as  $s > 0\%$  with 99% confidence); we refer to these clones as “adaptive,” and the clones falling outside the 99% confidence level as “neutral.” This 59% adaptive fraction is similar to the [Levy et al. \(2015\)](#) estimate of 50% adaptive lineages at generation 88.

### Whole-Genome Sequencing

To determine the genetic basis of adaptation we conducted whole-genome sequencing for 418 of the 3,883 unique barcoded clones with assigned fitness estimates (see [STAR Methods](#)). These included 333 adaptive clones, consisting of nearly every sampled clone with  $s > 5\%$  and many lower fitness clones



**Figure 2. Fitness Measurements Are Consistent across Replicates and Techniques**

(A) Comparison of fitness values for individual barcoded clones obtained from independent replicate assays conducted in the same experimental batch. (B and C) Comparison of fitness values for individual barcoded clones obtained from independent experimental batches (averaged over all replicates within a batch) of the fitness measurement assay. For (A)–(C), a small number of lineages with extreme fitness estimates in at least one replicate ( $s < -5\%$  or  $s > 20\%$ ) are not shown for increased resolution. (D–F) A comparison of our fitness measurements using the 4,800 clone pool and (D) fitness measurements from a 500 clone pool, (E) to their barcode lineage fitness measurements from the [Levy et al. \(2015\)](#) lineage tracking estimates, and (F) the pairwise fluorescence competition assay measurements, from [Levy et al. \(2015\)](#). Note that the lineages we classified as neutral but which were called adaptive by [Levy et al. \(2015\)](#), and vice versa, are highlighted by red boxes, with explanation in main text. The solid lines on all panels are  $Y = X$  while  $X$  and  $Y$  error bars show the fitness measurement errors (see [STAR Methods](#)). For each panel, we report the mean and SD of the difference in fitness for each comparison, grouped by low and high fitness clones. Systematic differences between measurements appear to be lower in low-fitness clones compared to high fitness clones, but the measurements are generally consistent throughout. We conducted extensive validation of our fitness estimation methodology, highlighted in [Figures S1, S2, S3, S4, and S5](#).

( $0\% < s < 5\%$ ). To understand the spectrum of neutral mutations, we also sequenced 85 neutral clones. Our sequenced clones thus covered the entire range of observed fitness values ([Figure S6](#), blue bars). We obtained  $\sim 20\times$  average and  $5\times$  minimum coverage for each clone. We called SNPs and short indels using a GATK-based pipeline and manual curation, and larger structural variants were identified with CLC Genomics Workbench (see [STAR Methods](#)). Sanger sequencing of 57 randomly chosen mutations that passed manual curation revealed no false positives (see [STAR Methods](#)). Across all clones (adaptive and neutral), we identified a total of 445 mutations ([Table 1](#); [Table S4](#); [Data S1](#)), including 352 point mutations, 44 insertion/deletion events, 4 chromosomal aneuploidy events, and 45 transposable element (TE) insertion events. A total of 211 clones (188 adaptive clones) have more than one mutation.

### Self-Diploidization Is an Adaptive Mechanism

In 83 adaptive clones, we observed the surprising presence of unambiguous heterozygous mutations, suggesting that many of the clones were diploid. To validate this, and to measure the frequency of diploidy, we developed a high throughput method to determine the ploidy of all 4,800 sampled clones, based on [Upshall et al. \(1977\)](#) (see [STAR Methods](#)). This method takes advantage of the stronger growth inhibition at  $25^\circ\text{C}$  of diploid cells compared to haploid cells in media containing benomyl; our assay was 99% concordant with flow cytometry ploidy analysis of a sample of  $\sim 800$  clones. Of the 4,800 clones, 43% from evolution E1 and 60% from evolution E2 were diploid ([Table S1](#)). We also performed mating assays (see [STAR Methods](#)) for  $\sim 1,200$  randomly chosen clones, including haploids and diploids from both E1 and E2, and found that every clone behaved as a

**Table 1. Summary of Mutations Observed in this Study**

Class	Locus	No. of Clones	No. of Clones with No Mutations Outside Locus	Synonymous	Missense	Nonsense	Coding In/Del	Coding TE Insertion	Noncoding SNV	Noncoding In/Del	Noncoding TE Insertion	Chromosome Amplification
<b>Neutral Diploid</b>												
	All	3	1	1	1		1					
<b>Neutral Haploid</b>												
	All	82	34	6	25		1	8	20	3	11	
<b>Adaptive Diploid</b>												
	All	237	102	31	77	8	4	3	43	7	5	4
<b>Adaptive Haploid with Nutrient Response Pathway Mutations</b>												
	<i>IRA1</i>	32	14		9	11	11	1				
	<i>IRA2</i>	12	4		9	2	1					
	<i>GPB1</i>	4	1		1	2	1					
	<i>GPB2</i>	15	8		2	8	5					
	<i>PDE2</i>	11	5		2	3	5	1				
	<i>RAS2</i>	1			1							
	<i>TFS1</i>	1			1							
	<i>CYR1</i>	3	2		3							
	<i>TOR1</i>	1	1		1							
	<i>KOG1</i>	1	1		1							
	<i>SCH9</i>	1			1							
	all other loci			14	33	3	2	2	15	1	11	
	All	82		14	64	29	25	4	15	1	11	
<b>Adaptive Haploid without Nutrient Response Pathway Mutations</b>												
	All	14	3	2	8	2	2	1	6		2	
<b>All Adaptive Haploid (Sum of Above Two Classes)</b>												
	All	96	3	16	72	31	27	5	21	1	13	

Mutations are tabulated by gene in different subsets of clones. “In/Del” stands for short “insertion/deletion” events, while “TE” stands for “transposable element.” The columns labeled “No. of Clones” and “No. of Clones with no mutations outside locus” should be interpreted as, e.g., three adaptive haploid clones with *CYR1* mutations, two of which have no mutations outside *CYR1*, or 237 adaptive diploid clones, 102 of which have no mutations aside from diploidy. See [Table S4](#) and [Data S1](#) for details for all the mutations used in this study. See [Figure S7](#) for validation that the diploidy mutations arose early in the evolution experiment.

*MAT $\alpha$*  strain (the mating type of the founding ancestor). Thus, all of the diploids apparently arose via self-diploidization to generate *MAT $\alpha$ /MAT $\alpha$*  diploids, rather than by mating type switching and subsequent mating between haploids of opposite mating types. Such self-diploidization has been observed to be beneficial in a prior glucose-limited evolution experiment (Gerstein et al., 2006).

Of our whole-genome sequenced clones, 240 were diploid, of which the vast majority—237 (99%)—were measured as adaptive, with an average fitness benefit of  $3.6\% \pm 0.6\%$ . This included 12 clones used for the pairwise competition assays in Levy et al. (2015), which had an average fitness benefit of 3.5% in that assay, validating diploidy as an adaptive mutation. Aside from three diploid clones carrying an extra copy of chromosome 11 (discussed below), there was no significant difference in the fitness of adaptive diploid clones that contained no additional mutations ( $n = 102$ ), as compared to either diploids with additional mutations that do not alter protein sequence ( $n = 53$ ), or diploids containing additional mutations (i.e., missense, nonsense, and insertion/deletion) that do alter protein sequence ( $n = 79$ ) ( $3.4\%$  versus  $3.2\%$  versus  $4.2\%$ ;  $p > 0.1$ ; ANOVA). This strongly suggests that diploidy is the only driving adaptive mutation in most or all of these clones. Three of the sequenced adaptive diploid clones contained an extra copy of chromosome 11, which conferred a significant fitness advantage beyond diploidy alone ( $s = 7.6\% \pm 0.6\%$ ;  $p \leq 0.0001$ ; ANOVA test for each of the four batches of fitness measurements). One additional diploid clone contained an extra copy of chromosome 12, but was not significantly more fit than the average diploid ( $s = 4.6\%$ ,  $p > 0.1$ ).

Of the 1,649 lineages that we determined to be diploids, 451 (27%) had been previously determined by the lineage tracking analysis of Levy et al. (2015)—without any knowledge of ploidy—to be lineages that were adaptive, with roughly the same fitness values, across both replicate evolution experiments. This suggests that many of these lineages were already self-diploidized by the time they were present in the barcoded population used to found the replicate evolutions; potentially the self-diploidization occurred during the transformation process itself when the barcodes were introduced into the cells. To investigate this, we measured the frequency of diploids throughout the Levy et al. (2015) replicate evolutions, and determined that at time zero the frequency of diploidy was low ( $\sim 1\%$ ; Figure S7). We also conducted additional 200-generation evolution experiments using the experimental conditions of Levy et al. (2015) but using an isogenic non-barcoded haploid ancestral population (i.e., that had not undergone transformation) and found that  $<0.1\%$  of sampled clones were diploid at generation 88, indicating that spontaneous self-diploidization under our adaptive growth conditions is a rare event. The possibility of transformation-induced diploidy prevents us from accurately estimating a mutation rate for self-diploidization, but it is clear that whole-genome duplication alone is beneficial under our growth conditions with a fitness effect of  $\sim 3.4\%$ .

#### Adaptive Haploid Clones Typically Carry a Single Adaptive Mutation

Of the 418 clones we sequenced, 178 were haploid, of which 96 were adaptive and 82 neutral. We found a significant excess in

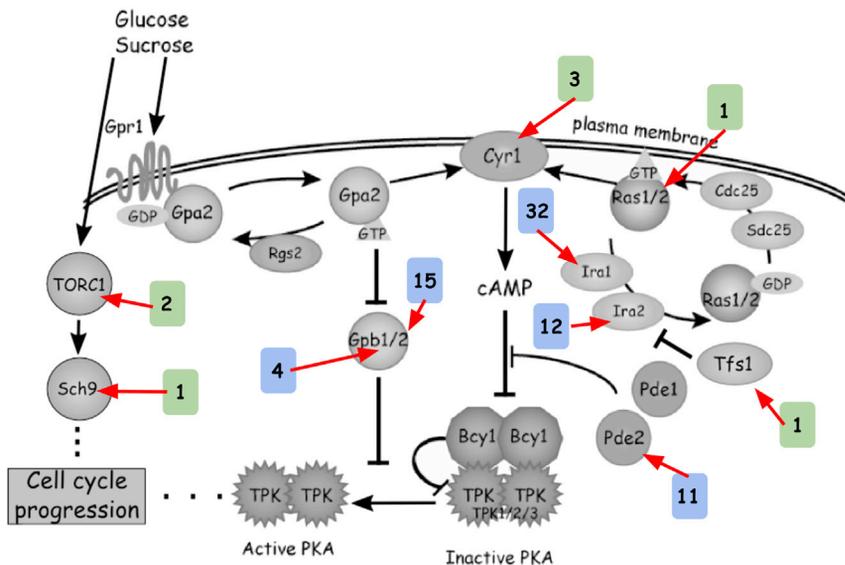
the total number of mutations in adaptive haploid clones compared to neutral haploid clones (1.95 versus 0.94 mutations per clone;  $p = 0.00004$ ; ANOVA; Table 1); note, the observed number of mutations in neutral clones (0.94 per clone) is higher than the expected 0.5 events per clone after 88 generations, based on the mutation rate estimates of Levy et al. (2015). The source of this excess is unknown, although it is possible that mutations may have been induced by transformation of the DNA barcodes. It has been speculated that transformation is mutagenic (Gaeveer et al., 2002; Shortle et al., 1984) and would be consistent with the transformation-induced diploidy hypothesized above.

The adaptive clones have, on average, almost exactly one additional mutation compared to neutral clones, suggesting that they indeed carry only a single adaptive mutation. The adaptive haploid clones also have a significantly larger proportion of protein sequence altering mutations (i.e., missense, nonsense, or insertion/deletion mutations) (73%) when compared to the neutral clones (46%) (Table 1;  $p = 0.0001$ , Fisher's exact test), strongly suggesting that the additional mutations in the adaptive clones impact protein function.

#### Adaptive Haploids Are Enriched for Mutations in the Nutrient Response Pathways

A hallmark of adaptive mutations in laboratory evolution experiments is the finding of recurrent mutations within genes or pathways, which is unlikely under neutral evolution. We define candidate adaptive targets as those loci with at least two independent adaptive mutations among our sequenced clones. None of the protein-altering mutations found in the neutral clones occurred in the same gene; by contrast, 77 of the 135 (57%) protein-altering mutations in the adaptive clones were found in recurrently mutated genes ( $p = 10^{-11}$ , Fisher's exact test). All of these 77 mutations were found in clones with different barcodes and are thus independent. The recurrent mutations in the adaptive clones occurred in six genes (*IRA1*, *IRA2*, *GPB1*, *GPB2*, *PDE2*, and *CYR1*), all of which are in the Ras/PKA pathway and are known to regulate yeast cell growth in response to glucose availability (reviewed in Conrad et al., 2014). A number of identical mutations occurred independently more than once: single mutations in *CYR1*, *GPB1*, and *GPB2* and two different mutations in *IRA1* each occurred twice independently, while a single mutation in *PDE2* occurred independently four times. Mutations in this pathway have been identified as adaptive in previous glucose-limited yeast evolution experiments (e.g., Kao and Sherlock, 2008; Wenger et al., 2011; reviewed in Long et al., 2015), with selective effects of  $\sim 10\%$ – $25\%$  per generation in chemostats. We also observed one mutation in each of three different genes belonging to the TOR/Sch9 pathway (*TOR1*, *KOG1*, and *SCH9*), which also integrates nutrient availability information with growth. We did not observe recurrent mutations in any other genes or pathways.

A total of 82 of our 96 (85%) sequenced adaptive haploid clones contained a mutation in either the Ras/PKA or TOR/Sch9 pathways (Figure 3; Table 1); 36 of these 82 clones had no other identified mutations, strongly indicating for these clones (and implying for the other clones) that the mutation in the Ras/PKA or TOR/Sch9 pathway gene is the causal adaptive mutation.



**Figure 3. Schematic of the Ras/PKA and TOR/Sch9 Pathways in Yeast and the Number of Adaptive Mutations Found per Gene**

The colored boxes denote the number of independent haploid lineages observed in our dataset with mutations in a particular gene. Blue boxes indicate mutations in negative regulators of cell-cycle progression, while green boxes indicate mutations in positive regulators. Modified from Figure S1 of Kao and Sherlock (2008).

We found that most diploid clones have a fitness advantage close to the mean for diploids without other mutations (~3.4%) with variations consistent with counting noise (Figure S3), again suggesting that these clones have functionally identical adaptive mutations—that is, solely diploidy. By contrast, lineages with mutations in the Ras/PKA and TOR/Sch9 nutrient response pathways have

We also note that four diploid clones (not included in the 82 described above) also carried mutations in the nutrient response pathway genes. Of the remaining adaptive haploid clones that did not have mutations in the Ras/PKA or TOR/Sch9 pathways, three were clones for which we were unable to identify any mutations, and 11 had mutations that did not appear to affect other nutrient response pathways (Table 2). We do not find any evidence for adaptive copy-number changes in any of our haploid clones.

In genes known to be positive regulators of the Ras/PKA and TOR/Sch9 pathways (*RAS2*, *CYR1*, *TOR1*, *KOG1*, *SCH9*, and *TFS1*) we identified only missense mutations, and for each of these genes there were only 1 to 3 clones with such mutations (Table 1). By contrast, in genes encoding negative regulators of the Ras/PKA pathway (*IRA1*, *IRA2*, *GPB1*, *GPB2*, and *PDE2*) many of the mutations were likely inactivating (insertion/deletion and nonsense) and mutations in these genes were observed much more frequently, with 4 to 32 mutant clones per gene (Table 1). These results suggest that most adaptive mutations in the positive regulator genes increase or modify activity (hypermorphic) and thus have a small mutational target size, while those in negative regulator genes of the nutrient response pathway decrease or abolish activity (hypomorphic). As expected of clones with hypermorphic mutations in the TOR/Sch9 pathway, those clones had increased rapamycin resistance (data not shown).

### The Fitness Effect of a Mutation Is Dependent on the Gene and the Mutation Type

We integrated our genotype data with our fitness estimates to study the distribution of fitness effects for all of our major mutation classes, generating a genotype-to-fitness map for the initial driver mutations in our evolution experiment (Figure 4). As the fitness benefits may not necessarily be gained during exponential growth, we also provide an additional y axis on the plot, showing the fitness per growth cycle (a factor of eight larger).

fitness benefits ranging from 5% to 15%, depending on the gene and type of mutation, suggesting a lack of functional equivalency between different adaptive mutations within these nutrient response pathway genes. Together, the diploidy ( $s \sim 3.4\%$ ) and nutrient response pathway mutations ( $s \sim 5\%–15\%$ ) explain the two major fitness classes observed in Levy et al. (2015) (see Figure 3B of that work) and in our fitness measurement assays (see Figure S6).

We conducted a number of ANOVA tests for the effects of gene identity, mutation type, and the presence of additional coding mutations on the fitness of our clones containing nutrient response pathway mutations. We found significant effects of both gene identity ( $p < 10^{-7}$ ; ANOVA), and mutation type ( $p < 10^{-3}$ ; ANOVA after controlling for gene effects for three of four batches) on the fitness of these lineages. These differences can even be found between paralogs: the 32 mutations in *IRA1* confer a significantly greater fitness advantage, on average, than the 12 mutations in its paralog *IRA2* (12.9% versus 10.2%) ( $p < 0.05$ ; ANOVA), and mutations in *GPB2* confer a significantly greater fitness advantage than mutations in *GPB1* (10.4% versus 6.2%) ( $p < 10^{-4}$ ; ANOVA). In addition, missense mutations in *IRA1* confer a significantly lower fitness benefit than nonsense or insertion/deletion mutations within the same gene ( $p \leq 0.05$ , ANOVA for three of four batches).

The fitness distribution for lineages carrying mutations in *GPB2* is remarkably narrow within replicates (SD <1% per generation across all replicates), particularly when compared to other nutrient response pathway genes such as *IRA1* (SD of 1%–3% per generation). Note, this variation in *GPB2* is substantially less than the average variation observed between replicates and batches for high fitness lineages (Figure 2). One possible explanation is that every mutation in *GPB2* completely abolishes gene function; alternatively, partial loss of *GPB2* function may still lead to the same level of Ras/PKA pathway activation as a complete loss of function, resulting in these highly consistent fitness estimates. In either case, the lack of fitness

**Table 2. Mutations in Adaptive Haploid Clones without a Nutrient Response Pathway Mutation**

Lineage ID	Fitness (%)	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
7538	8.2	<i>ERG1</i> ; missense; H220Y	<i>THI3</i> ; nonsense; S18*			
7953	10.0					
13183	13.9					
14688	2.3	<i>STE3</i> ; missense; I141T	<i>RXT2</i> ; missense; N63Y			
18152	13.9	<i>KT112</i> ; missense; K208Q				
21863	3.2					
26598	4.0	YLR157W-E; TE insertion				
53054	1.6	<i>ATG17</i> ; upstream SNV; 2655 bp A/G				
60700	8.5	<i>SSK2</i> ; nonsense; E702*				
88494	6.4	<i>NCL1</i> ; upstream SNV; 374 bp T/A				
225103	2.1	YKL068W-A upstream TE insertion	<i>IES3</i> ; missense; N241T	<i>LAA1</i> ; synonymous; T623T	<i>SEC4</i> ; synonymous; N65N	
254044	2.2	<i>TCB2</i> ; upstream SNV; 99 bp C/A	<i>BAT1</i> ; upstream SNV; 498 bp A/T	<i>LEU4</i> ; missense; I264V	<i>DMA1</i> ; frameshift; 989 bp	<i>SUP51</i> , <i>CYR1</i> upstream TE insertion
262917	2.6	<i>FPK1</i> ; frameshift; 2113bp	<i>POP4</i> ; upstream SNV; 246 bp C/G			
304483	3.1	YOL014W; missense; L65M	<i>BEM2</i> ; missense; D2054Y	<i>PAU16</i> ; upstream SNV; 835 bp AT/TG		

variation among the lineages with mutations in *GPB2* demonstrates the precision of our fitness estimates and further suggests that the fitness differences observed between replicates and batches (Figure 2) may be due to biological variation in fitness due to slight differences in conditions rather than estimation error.

We also tested for the presence of additional adaptive mutations in the adaptive haploid clones containing nutrient response pathway mutations. We found that the 32 clones with both a nutrient response pathway mutation and an additional protein sequence altering mutation do not have a significantly different fitness than the 50 clones with a nutrient response pathway mutation alone ( $p < 0.05$  for only one of the four batches; ANOVA controlling for gene and mutation type).

### Not Every Gene in the Ras/PKA Pathway Is a Target of Adaptation

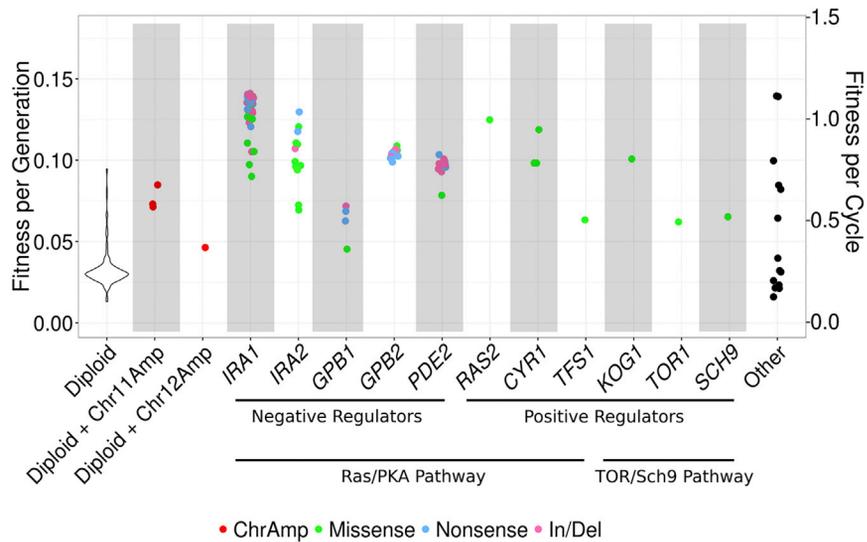
Among our sequenced adaptive clones, we found putative hypomorphic mutations in most of the negative regulators of the Ras/PKA pathway (*IRA1*, *IRA2*, *GPB1*, *GPB2*, and *PDE2*) but no mutations in *PDE1*. We hypothesized that *PDE1* mutations did not confer a substantial fitness advantage, as *Pde1* has a lower affinity for cAMP than *Pde2* (Londesborough and Lukkari, 1980). To test this hypothesis, and to confirm that loss of any of the five negative regulators of the Ras/PKA pathway we observe as mutated is indeed adaptive, we constructed whole-gene deletions of *IRA1*, *IRA2*, *GPB1*, *GPB2*, *PDE1*, and *PDE2*, as well as

the pseudogene YFR059C as a control, and assayed their fitness using fluorescence-based pairwise competition assays (see STAR Methods). As predicted, we found that the fitness of the *PDE1* deletion mutant was indistinguishable from neutrality, while deletion of the other genes was highly beneficial (Figure 5) with the fitness benefit roughly similar to that of the detected mutations in these genes.

### DISCUSSION

One of the key goals of the study of adaptive evolution is to characterize the molecular basis and fitness effects of a comprehensive set of adaptation-driving mutations. We have overcome several challenges to achieve this goal: sampling a large number of independent clones without any bias for the type of adaptive event (e.g., point mutation versus structural variant versus epigenetic change), identifying adaptive events across the whole genome, and estimating the fitness effects of each of these mutations in a high-throughput manner, with high confidence and at a low cost per assay (~\$0.07 per clone per replicate measurement). In addition, as exemplified by the small variation in the many independent fitness measurements for *GPB2* mutants, our fitness measurements are both sensitive and precise.

By sampling adaptive mutations while they are still collectively a modest fraction of the population, we were able to identify the two major (and perhaps only) classes of adaptive mutations that drive early evolution in our experiment: (1) self-diploidization



**Figure 4. The Fitness Spectrum (Genotype-to-Fitness Map) of Evolved Clones with Different Adaptive Mutations**

The inverse variance weighted fitness averaged across all batches and replicates is plotted. Mutations are colored by their molecular basis (i.e., chromosomal amplification, missense, nonsense, or insertion/deletion). The “other” class includes the 14 adaptive haploid clones for which we did not identify a nutrient response pathway mutation. Within-batch SDs (not shown for clarity) are  $\leq 1\%$  for  $>90\%$  of clones with nutrient response pathway mutations, while between-batch SDs are  $\sim 2\%$  for all clones. To highlight the effect of single mutations on fitness, the six diploid clones with nutrient response pathway mutations are not shown. We show per-cycle fitness (eight generations per cycle) as a secondary y axis (right side), as the fitness benefit of these mutations may not exclusively be due to changes in per-generation fermentative growth rate, but due to changes in other parts of the growth cycle such as growth lag, diauxic shift, aerobic growth, or increased viability after stationary phase. [Figure S6](#) shows the distribution of fitness effects of our 4,800 sampled and 418 sequenced clones.

( $s \sim 3.4\%$ ), and (2) mutations presumably activating the Ras/PKA or TOR/Sch9 pathways ( $s \sim 5\%–15\%$ ). These two classes of mutations explain the fitness advantages of 319/333 (96%) of our sequenced adaptive clones, suggesting that in our system early stage adaptation is driven by only a small number of mutational classes. We can also be certain that we did not miss a large class of difficult to identify adaptive events, such as mutations in repetitive regions, complex structural changes, or epigenetic modifications.

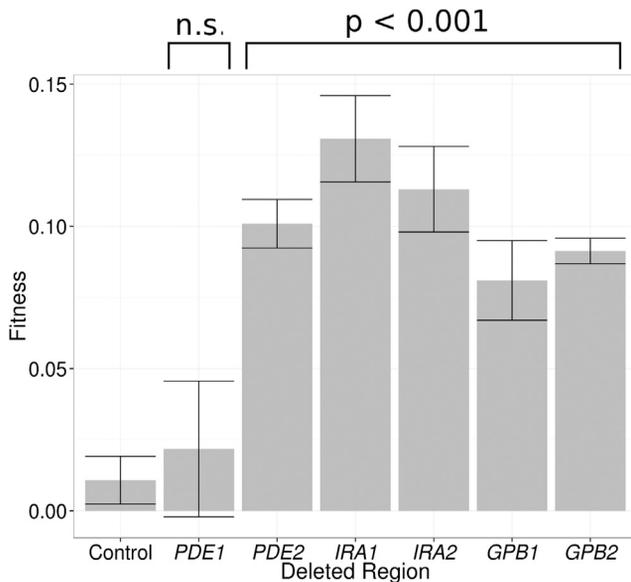
We found a large number of recurring large-effect adaptive mutations in a small number of genes. In one case only a single member of a paralog pair, *PDE2* and not *PDE1*, had any observed mutations. We confirmed that the reason we did not observe mutations in *PDE1* was not due to insufficient sampling depth, but rather was due to *PDE1* mutations not being adaptive under our experimental conditions. The results make us confident that we have generated a comprehensive map of the predominant adaptation-driving mutations in *S. cerevisiae* grown in one specific environment.

Note, we have not attempted to identify every potentially adaptive mutation in our experimental condition, rather we have identified most of the mutations that drive or are likely to drive the evolutionary dynamics of our system. In this system, with its well-mixed population, any adaptive mutation that is either too selectively weak or has a very low rate of occurrence cannot effectively drive the adaptive dynamics, because of clonal interference ([Levy et al., 2015](#)). For example, if the target sizes for adaptive mutations in two genes are  $k_1$  and  $k_2$  respectively, with selective advantages  $s_1$  and  $s_2$ , then after a time  $T$  in a large population the ratio of the fractions of the population of the two classes of mutants are  $k_1 \exp(s_1 T)$  and  $k_2 \exp(s_2 T)$ . If  $T = 88$  generations, as for our sampled clones, with  $s_1 - s_2 = 5\%$ , and the same target sizes ( $k_1 = k_2$ ), the mutant with 5% greater fitness benefit will be observed 100 times as often. However, the muta-

tional target size is also important: if  $k_2$  were  $100\times$  larger than  $k_1$  (e.g.,  $k_2$  includes many possible beneficial loss of function mutations while  $k_1$  includes only very few beneficial gain of function mutations), this compensates for the selective effect and mutations in the two genes will become comparable fractions of the population. Therefore, both selective advantage and the mutational target size are important in determining which mutations drive adaptive evolution.

The importance of both parameters may explain why we observed few candidate adaptive mutations in regulatory regions of Ras/PKA or TOR/Sch9 pathway genes. Indeed, we observed only one possible case, a transposon insertion upstream of the *CYR1* gene, in a clone for which there were no other obvious adaptive mutations. Such mutations may therefore be rarer and/or confer a smaller selective advantage than changes to the actual protein sequences in our system and experimental condition.

The first key mutational event that we identified here was self-diploidization. The presence of a diploid fitness advantage in our growth condition is consistent with previous work showing that self-diploidization frequently fixes in yeast populations evolving under glucose limitation ([Gerstein et al., 2006](#)), but contrasts with the fitness disadvantage of diploids relative to haploids found under glucose limitation ([Adams and Hansche, 1974](#); [Zeyl et al., 2003](#)) and no difference in fitness under nitrogen limitation ([Hong and Gresham, 2014](#)). Note, however, that these studies were performed in environmental conditions different from ours (chemostats versus batch culture), which could significantly modify the relative fitness of haploids and diploids. This is consistent with a prior study that has found that the relative growth rates of haploid and diploid cells is highly dependent on both the specific strain genotype and the environment ([Zörgö et al., 2013](#)). A large body of work (reviewed in [Otto, 2007](#)) has sought an explanation for the evolution of diploidy in eukaryotes



**Figure 5. Fitness of Clones with Synthetic Whole-Gene Deletions in Negative Regulators of the Ras/PKA Pathway**

Fitness was assayed using pairwise competitions against a fluorescently tagged ancestral clone. The plot shows the fitness for each of the constructed deletions and the error bars show SEM. p Values are for t tests comparing the fitness of each gene deletion to the control, which is a deletion of the pseudogene YFR059C. n.s., non-significant (meaning  $p > 0.05$ ).

and the frequent polyploidization events in the evolutionary history of many organisms, including *S. cerevisiae* (Marcet-Houben and Gabaldón, 2015). Our work and that of Zörgö et al. (2013) suggest that diploidy may arise under some conditions due to a direct fitness advantage for diploids when compared to isogenic haploids. Further work is needed to determine the physiological basis for this fitness advantage and the generality of this advantage in other conditions.

The second major type of adaptive event targeted genes in the Ras/PKA and TOR/Sch9 nutrient response pathways. Previous work has shown that mutations in these pathways exhibit strong pleiotropic effects. For example, natural genetic variation present in many genes in the Ras/PKA pathway responds to selection for growth at 40°C (Parts et al., 2011). In addition, loss of function mutations in the TOR/Sch9 pathway result in an increased replicative lifespan (number of viable cell divisions per cell) (Kaeberlein et al., 2005), as do mutants that decrease activity of the Ras/PKA pathway (Fabrizio et al., 2004; Lin et al., 2000). The study of the pleiotropic nature of fitness trade-offs (antagonistic pleiotropy), is critical to understanding adaptive evolution in the laboratory and in nature. Our DNA barcode-based approach allows for the isolation and economic measurement of the individual fitness values of large pools of mutants, which will be of great use in investigating such evolutionary trade-offs.

In summary, we have conducted an in-depth survey of the molecular nature and associated fitness effects of the adaptive mutations in an evolving system, generating a genotype-to-fitness map for the mutations that drive the initial adaptive evolution. This approach opens the possibility of a far more in-depth understanding of adaptive evolution by de novo mutations and gives

us a new way to assay the fitness landscapes in evolving systems comprehensively, economically, and precisely.

## STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Sampling Clones
  - Pairwise Fluorescence Competition Assay Measurements
  - Pooled-Clone Fitness Measurement Assay
  - Construction of a Strain with Ancestral Fitness
  - Optimizing Sequencing Costs
  - Pooling the 4,800 Clones Sampled at Generation 88
  - Conducting the Fitness Measurement Assay
  - DNA Extractions from Each Sample
  - PCR Amplification of the Barcode Locus
  - Removal of the Reference Strain Amplicons Using Restriction Digestion and Size Selection
  - Multiplexing and Amplicon Sequencing
  - Initial Processing of the Amplicon Sequencing Data
  - Whole-Genome Sequencing
  - Determination of Mating Type and Ploidy
  - Construction of Gene Deletions in the Ras/PKA Pathway
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Introduction to Fitness Estimation Methodology
  - Noise Model
  - Checks on the Noise Model
  - Fitness Assay
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources
  - Software

## SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, four tables, and one data file and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2016.08.002>.

## AUTHOR CONTRIBUTIONS

S.V., B.D., J.B., S.F.L., D.S.F., G.S., and D.A.P. conceived of the project and designed the experiments. S.V., B.D., Y.L., A.A., J.C., E.E., K.G.-S., L.H., and J.B. conducted the experiments and analyzed the data. S.V., B.D., G.S., and D.A.P. wrote the manuscript with substantial assistance from the other authors.

## ACKNOWLEDGMENTS

We wish to thank all members of the D.A.P., G.S., and D.S.F. labs for useful discussions, and Michael Desai, Sergey Kryazhimskiy, Katja Schwartz, Dave Yuan, and Jake Cherry for technical help. We thank the Stanford Shared FACS facility for use of their flow cytometers, and the Stanford Center for Personalized Genomics and Medicine and NextSEQ for Illumina sequencing services. S.V. is supported by NIH/NHGRI T32 HG000044 and the Stanford

Center for Computational, Human and Evolutionary Genomics (CEHG); E.E. by NSF GFRP DGE-1247312; J.C. by NSF GRFP DGE-114747; A.A. by a Stanford Bio-X Bowes Fellowship; L.H. by NIH grant R01 GM110275 and a fellowship from CEHG; J.B. and S.F.L. by the Louis and Beatrice Laufer Center; and D.S.F. by NSF PHY-1305433 and NIH R01 HG003328. The work was supported by NIH grants R01 HG003328 and GM110275 to G.S. and R01 GM115919, GM10036601, and GM097415 to D.A.P. Data were collected on an instrument in the Shared FACS Facility obtained using NIH S10 Shared Instrument grant RR027431.

Received: March 7, 2016

Revised: June 7, 2016

Accepted: July 29, 2016

Published: September 1, 2016

## REFERENCES

- Adams, J., and Hansche, P.E. (1974). Population studies in microorganisms. I. Evolution of diploidy in *Saccharomyces cerevisiae*. *Genetics* 76, 327–338.
- Bank, C., Hietpas, R.T., Jensen, J.D., and Bolon, D.N.A. (2015). A systematic survey of an intragenic epistatic landscape. *Mol. Biol. Evol.* 32, 229–238.
- Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461, 1243–1247.
- Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580.
- Bozek, K., Wei, Y., Yan, Z., Liu, X., Xiong, J., Sugimoto, M., Tomita, M., Pääbo, S., Pieszek, R., Sherwood, C.C., et al. (2014). Exceptional evolutionary divergence of human muscle and brain metabolomes parallels human cognitive and physical uniqueness. *PLoS Biol.* 12, e1001871.
- Conrad, M., Schothorst, J., Kankipati, H.N., Van Zeebroeck, G., Rubio-Teixeira, M., and Thevelein, J.M. (2014). Nutrient sensing and signaling in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* 38, 254–299.
- Cousin, A., Heel, K., Cowling, W.A., and Nelson, M.N. (2009). An efficient high-throughput flow cytometric method for estimating DNA ploidy level in plants. *Cytometry A* 75, 1015–1019.
- Darwin, C. (1872). *The Origin of Species* (London: John Murray).
- Davies, J., and Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol. Mol. Biol. Rev.* 74, 417–433.
- De Meester, L., Gomez, A., Okamura, B., and Schwenk, K. (2002). The monopolization hypothesis and the dispersal-gene flow paradox in aquatic organisms. *Acta Oecol.* 23, 121–135.
- Desai, M.M., and Fisher, D.S. (2007). Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* 176, 1759–1798.
- Dettman, J.R., Rodrigue, N., Melnyk, A.H., Wong, A., Bailey, S.F., and Kassen, R. (2012). Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Mol. Ecol.* 21, 2058–2077.
- Fabrizio, P., Pletcher, S.D., Minois, N., Vaupel, J.W., and Longo, V.D. (2004). Chronological aging-independent replicative life span regulation by *Msn2/Msn4* and *Sod2* in *Saccharomyces cerevisiae*. *FEBS Lett.* 557, 136–142.
- Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. *Nat. Methods* 11, 801–807.
- Gerstein, A.C., Chun, H.-J.E., Grant, A., and Otto, S.P. (2006). Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS Genet.* 2, e145.
- Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391.
- Gietz, R.D., and Woods, R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* 350, 87–96.
- Givnish, T.J. (2015). Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. *New Phytol.* 207, 297–303.
- Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. *Nature* 481, 306–313.
- Gresham, D., Desai, M.M., Tucker, C.M., Jenq, H.T., Pai, D.A., Ward, A., DeSevo, C.G., Botstein, D., and Dunham, M.J. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* 4, e1000303.
- Herron, M.D., and Doebeli, M. (2013). Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biol.* 11, e1001490.
- Hietpas, R.T., Bank, C., Jensen, J.D., and Bolon, D.N.A. (2013). Shifting fitness landscapes in response to altered environments. *Evolution* 67, 3512–3522.
- Hong, J., and Gresham, D. (2014). Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet.* 10, e1004041.
- Kaeberlein, M., Powers, R.W., 3rd, Steffen, K.K., Westman, E.A., Hu, D., Dang, N., Kerr, E.O., Kirkland, K.T., Fields, S., and Kennedy, B.K. (2005). Regulation of yeast replicative life span by *TOR* and *Sch9* in response to nutrients. *Science* 310, 1193–1196.
- Kao, K.C., and Sherlock, G. (2008). Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nat. Genet.* 40, 1499–1504.
- Korolev, K.S., Xavier, J.B., and Gore, J. (2014). Turning ecology and evolution against cancer. *Nat. Rev. Cancer* 14, 371–380.
- Kryazhimskiy, S., Rice, D.P., Jerison, E.R., and Desai, M.M. (2014). Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344, 1519–1522.
- Kvitek, D.J., and Sherlock, G. (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* 7, e1002056.
- Kvitek, D.J., and Sherlock, G. (2013). Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genet.* 9, e1003972.
- Lachance, J., and Tishkoff, S.A. (2013). Population genomics of human adaptation. *Annu. Rev. Ecol. Evol. Syst.* 44, 123–143.
- Landau, D.A., Carter, S.L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M.S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 152, 714–726.
- Lang, G.I., Rice, D.P., Hickman, M.J., Sodergren, E., Weinstock, G.M., Botstein, D., and Desai, M.M. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500, 571–574.
- Levy, S.F., Blundell, J.R., Venkataram, S., Petrov, D.A., Fisher, D.S., and Sherlock, G. (2015). Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519, 181–186.
- Lin, S.J., Defossez, P.A., and Guarente, L. (2000). Requirement of NAD and *SIR2* for life-span extension by calorie restriction in *Saccharomyces cerevisiae*. *Science* 289, 2126–2128.
- Londesborough, J., and Lukkari, T.M. (1980). The pH and temperature dependence of the activity of the high Km cyclic nucleotide phosphodiesterase of bakers’ yeast. *J. Biol. Chem.* 255, 9262–9267.
- Long, A., Liti, G., Luptak, A., and Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat. Rev. Genet.* 16, 567–582.
- Marcet-Houben, M., and Gabaldón, T. (2015). Beyond the whole-genome duplication: phylogenetic evidence for an ancient interspecies hybridization in the baker’s yeast lineage. *PLoS Biol.* 13, e1002220.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- Nowell, P.C. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23–28.

- Oleksyk, T.K., Smith, M.W., and O'Brien, S.J. (2010). Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 185–205.
- Otto, S.P. (2007). The evolutionary consequences of polyploidy. *Cell* 131, 452–462.
- Palmer, A.C., and Kishony, R. (2013). Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat. Rev. Genet.* 14, 243–248.
- Parts, L., Cubillos, F.A., Warringer, J., Jain, K., Salinas, F., Bumpstead, S.J., Molin, M., Zia, A., Simpson, J.T., Quail, M.A., et al. (2011). Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* 21, 1131–1138.
- Pennings, P.S. (2012). Standing genetic variation and the evolution of drug resistance in HIV. *PLoS Comput. Biol.* 8, e1002527.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rich, M.S., Payen, C., Rubin, A.F., Ong, G.T., Sanchez, M.R., Yachie, N., Dunham, M.J., and Fields, S. (2016). Comprehensive analysis of the *SUL1* promoter of *Saccharomyces cerevisiae*. *Genetics* 203, 191–202.
- Shortle, D., Novick, P., and Botstein, D. (1984). Construction and genetic characterization of temperature-sensitive mutant alleles of the yeast actin gene. *Proc. Natl. Acad. Sci. USA* 81, 4889–4893.
- Sliwa, P., and Korona, R. (2005). Loss of dispensable genes is not adaptive in yeast. *Proc. Natl. Acad. Sci. USA* 102, 17670–17674.
- Soulebeau, A., Aubriot, X., Gaudeul, M., Rouhan, G., Hennequin, S., Haevermans, T., Dubuisson, J., and Jabbour, F. (2015). The hypothesis of adaptive radiation in evolutionary biology: hard facts about a hazy concept. *Org. Divers. Evol.* 15, 747–761.
- Stinchcombe, J.R., and Hoekstra, H.E. (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100, 158–170.
- Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D., and Gaut, B.S. (2012). The molecular diversity of adaptive convergence. *Science* 335, 457–461.
- Toprak, E., Veres, A., Michel, J.-B., Chait, R., Hartl, D.L., and Kishony, R. (2011). Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* 44, 101–105.
- Upshall, A., Giddings, B., and Mortimore, I.D. (1977). The Use of Benlate for Distinguishing Between Haploid and Diploid Strains of *Aspergillus nidulans* and *Aspergillus terreus*. *J. Gen. Microbiol.* 100, 413–418.
- Vitti, J.J., Grossman, S.R., and Sabeti, P.C. (2013). Detecting natural selection in genomic data. *Annu. Rev. Genet.* 47, 97–120.
- Voordeckers, K., and Verstrepen, K.J. (2015). Experimental evolution of the model eukaryote *Saccharomyces cerevisiae* yields insight into the molecular mechanisms underlying adaptation. *Curr. Opin. Microbiol.* 28, 1–9.
- Warringer, J., Zörgö, E., Cubillos, F.A., Zia, A., Gjuvsland, A., Simpson, J.T., Forsmark, A., Durbin, R., Omholt, S.W., Louis, E.J., et al. (2011). Trait variation in yeast is defined by population history. *PLoS Genet.* 7, e1002111.
- Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312, 111–114.
- Wenger, J.W., Piotrowski, J., Nagarajan, S., Chiotti, K., Sherlock, G., and Rosenzweig, F. (2011). Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLoS Genet.* 7, e1002202.
- Zeyl, C., Vanderford, T., and Carter, M. (2003). An evolutionary advantage of haploidy in large yeast populations. *Science* 299, 555–558.
- Zörgö, E., Chwialkowska, K., Gjuvsland, A.B., Garré, E., Sunnerhagen, P., Liti, G., Blomberg, A., Omholt, S.W., and Warringer, J. (2013). Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genet.* 9, e1003388.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
ExoSap-It	Affymetrix	Cat#78201
OneTaq 2X Master Mix with Standard Buffer	New England Biolabs	Cat#M0482L
10XCutSmart buffer	New England Biolabs	Cat#B7204S
ApaI restriction enzyme	New England Biolabs	Cat#R0507S
<b>Critical Commercial Assays</b>		
Zymo YeaStar Genomic DNA kit	Zymo Research	Cat#D2002
BioBasic 96 yeast genomic DNA extraction kit	BioBasic	Cat#BS8357
Nextera TD buffer and TDE1 enzyme	Illumina	Cat#FC-121-1030
KAPA HiFi Library Amplification Kit	KAPA Biosystems	Cat#KK2612
Nextera Index kit	Illumina	Cat#FC-121-1012
Agencourt AMPure XP magnetic beads	Beckman Coulter	Cat#A63880
QIAquick PCR purification kit	QIAGEN	Cat#28106
Qubit HS DNA quantitation kits	ThermoFisher	Cat#Q-33120
<b>Deposited Data</b>		
All Illumina sequencing data for both the whole-genome sequencing and the fitness measurement assays	This paper	NIH BioProject: PRJNA310010
<i>S. cerevisiae</i> (strain S288C) reference genome, version R64-1-1	<i>Saccharomyces</i> Genome Database (SGD)	<a href="http://www.yeastgenome.org">www.yeastgenome.org</a>
<i>S. cerevisiae</i> reference genome that includes the DNA barcode locus: <i>sacCer3</i> S288C	Levy et al., 2015	NIH BioProject: PRJNA310010
<b>Experimental Models: Organisms/Strains</b>		
<i>S. cerevisiae</i> YFP-tagged ancestor strain	Levy et al., 2015	GSY5306
<i>S. cerevisiae</i> neutral strain with ApaI restriction site in barcode, used for pooled fitness assay	This paper	GSY5929
<b>Sequence-Based Reagents</b>		
Forward Primer to amplify barcode region for Sanger sequencing (PS1): CCCGCAGAGTACTGCAATTT	This paper	N/A
Reverse Primer to amplify barcode region for Sanger sequencing (PS2): TGCACGAAAAGCAAACAAC	This paper	N/A
Primers to knock out Ras/PKA negative regulator genes: see <a href="#">STAR Methods</a>	This paper	N/A
Primers for PCR amplification of the barcode locus: see <a href="#">STAR Methods</a>	This paper	N/A
<b>Software and Algorithms</b>		
TrimGalore, version 0.3.7	Felix Krueger, Babraham Bioinformatics	<a href="http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/">http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/</a>
Novoalign, version 3.02.02	Novocraft Technologies	<a href="http://www.novocraft.com/products/novoalign/">http://www.novocraft.com/products/novoalign/</a>
PicardTools version 1.105(1632)	Broad Institute	<a href="http://broadinstitute.github.io/picard">http://broadinstitute.github.io/picard</a>
GATK version 3.2.2	McKenna et al. (2010)	N/A
Bedtools v2.17.0	Quinlan and Hall (2010)	N/A
Tandem Repeat Finder	Benson (1999)	N/A
CLC Genomics Workbench version 8.5	QIAGEN	<a href="http://www.clcbio.com">www.clcbio.com</a>
Pipeline to determine the number of barcode reads	This paper	<a href="https://github.com/sunthedeep/BarcodeCounter">https://github.com/sunthedeep/BarcodeCounter</a>

(Continued on next page)

### Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
96 well filter plates	Pall Life Sciences	Cat#8039
500mL DeLong flasks	Bellco	Cat#2510-00500
2mL yellow phase lock tube	5 PRIME	Cat#2302830
E-Gel SizeSelect agarose gels	ThermoFisher	Cat#G661002

### CONTACT FOR REAGENT AND RESOURCE SHARING

Further information regarding the manuscript may be directed to the lead contact Dmitri Petrov ([dpetrov@stanford.edu](mailto:dpetrov@stanford.edu)). Requests for reagents may be directed to, and will be fulfilled by Gavin Sherlock ([gsherloc@stanford.edu](mailto:gsherloc@stanford.edu)).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

The yeast strains used in this study can be grown and maintained using standard methods (e.g., YPD media in test tubes, glycerol stocks for long term storage at  $-80^{\circ}\text{C}$ ), but should be propagated in the selection environment (glucose limited minimal media) for optimal phenotypic and fitness measurements.

### METHOD DETAILS

#### Sampling Clones

Evolved yeast clones were isolated by plating for single colonies from frozen samples of the generation-88 time point of a previously reported serial batch transfer evolution experiment seeded by a population of individually genome-barcoded yeast cells (500,000 barcodes total); each batch cycle consisted of 8 generations of growth with glucose as the known limiting nutrient at an initial concentration of 1.5% ([Levy et al., 2015](#)). We selected 3,840 colonies from replicate experiment E1 and 960 colonies from replicate E2 for a total of 4,800 individual evolved clones. A portion of each colony was resuspended in 20% glycerol in 96 well plates and immediately frozen at  $-80^{\circ}\text{C}$ ; the remaining portion was used to identify the barcode residing in that clones genome. To identify the barcode, we amplified the genomic region carrying the barcode with PCR using the following primers:

PS1 - CCCGCAGAGTACTGCAATTT  
PS2 - TGCACGAAAAGCAAACAAC

The PCR products were purified using ExoSap-It (Affymetrix # 78200) and sequenced by Sanger technology, using PS2 as the sequencing primer, and then identified from among the set of 500,000 barcodes described in [Levy et al. \(2015\)](#).

#### Pairwise Fluorescence Competition Assay Measurements

Fluorescence-based fitness assays were conducted as in [Levy et al. \(2015\)](#). Briefly, the individual clone to be assayed was grown in liquid culture and then mixed with a YFP-tagged ancestral clone in a 1:9 ratio. This mixture was sampled over 32 generations (four 8-generation batch cycles) in conditions identical to the initial evolution experiments of [Levy et al. \(2015\)](#). The relative frequencies of the sample and the ancestor were estimated at each time-point using flow cytometry at the Stanford Shared FACS facility. An exponential model was then fit to these data to estimate fitness.

#### Pooled-Clone Fitness Measurement Assay

##### Overview

All 4,800 isolated clones were pooled into a single culture, then mixed with a clone with ancestral fitness in a 1:9 ratio and competed by culturing the mixture under conditions identical to the initial evolution for 32 generations (four 8-generation cycles), with samples being stored at every transfer. Barcode frequencies were tracked using Illumina HiSeq technology, and fitness was estimated using these frequency trajectories. We performed the fitness measurement assay a total of 11 times, in four independent batches, each time with two or three replicate flasks (see below for complete details).

##### Design of the Fitness Measurement Assay

The goal of the fitness measurement assay is to cheaply, easily, and accurately measure the fitness of many barcoded clones in parallel; any desired set of clones (as long as each clone contains a unique barcode at the same genomic position) can be pooled and used in this assay. In this protocol, we competed a pool of 4,800 clones sampled from generation 88 of the evolution experiments of [Levy et al. \(2015\)](#) against a clone with ancestral fitness, for a period of 32 generations of batch culture competition. We then estimated

the frequencies of the clones at five time points by Illumina amplicon sequencing of their DNA barcodes. The frequencies from four of these five time points (described in section 3.4) were then used to estimate fitness ( $s$ ). We used the same growth conditions as the evolution experiments of [Levy et al. \(2015\)](#), where the population goes through a bottleneck of  $\approx 5 \cdot 10^7$  cells every 48 hr, i.e., at the time it is transferred to fresh media ( $\approx 8$  generations between transfers).

To accurately measure the fitness of many clones, each adaptive lineage must be at a large enough frequency such that stochastic effects at the bottleneck process are mitigated. In addition, most of the initial population must consist of the ancestral genotype, so that fitness is measured in a condition dominated by the ancestor. To fulfill all of these criteria we first pooled 4,800 sampled clones, then seeded this pool at an initial frequency of 10% in the population and competed against an ancestral strain that made up  $\approx 90\%$  of the initial population. Each sampled clone thus had a population size in the bottleneck of  $\approx 0.1 \cdot 5 \cdot 10^7 / 5000 = 1,000$  cells. However, there are biological fluctuations from stochasticity in the lag time before new growth after dilution, in birth/death fluctuations near the bottleneck, and in the sampling induced by the dilution process itself. These give rise to fluctuations in the bottleneck population from one cycle to the next of  $\pm \sqrt{\beta n}$ . In our experiments, we estimated  $\beta \approx 10$  (see: Quantification and Statistical Analysis) so that these effects are relatively small. Furthermore, for beneficial mutations with  $s > 1\%$ , the systematic increases in population due to selection are larger than the stochastic fluctuations. Thus the stochastic effects are relatively small for most adaptive lineages in our fitness assay.

The large population size and the short time for the assay (32 total generations) also ensured that any new adaptive mutations that arose during the course of the assay had no significant impact on the fitness estimates of any single clone. The fraction of lineages that will get taken over by new mutations with a particular range of  $s$  can be approximately bounded by  $\mu/s \cdot e^{sT}$ , for a representative  $s$ . In our fitness assays, the total adaptive mutation rate per generation  $\mu$  for mutations with  $s > 5\%$  is  $\approx 10^6$  ([Levy et al., 2015](#)), highly adaptive mutations have  $s \sim 10\%$ , and the assay is conducted for  $T = 32$  generations. This gives us an upper bound of  $10^4$  for the expected fraction of lineages dominated by a mutant that came up during the evolution. The effect of new adaptive mutations is thus negligible in our fitness assays.

Ideally, the fraction of the population that consists of mutant clones would remain a small fraction of the total population throughout the 32 generations of growth in these assays. The dynamics of the limiting resource depend on the physiology of the dominant type(s) in the population, and if a non-ancestral type dominated it could change those dynamics in a way which affected different mutants differently. Additionally, if the ancestral clone gets to a low frequency, it becomes challenging to estimate fitness of a mutant relative to the ancestral type. As we describe in [Figure S1](#), the mutant clones started at a higher frequency within the barcoded class than planned and reached substantial frequency in the total pool at late time points. This might be one of the sources of systematic variations of measured fitnesses between experiments that we observe. Future experiments would be well served to minimize the effect of large populations of mutants changing the environment.

### Construction of a Strain with Ancestral Fitness

We realized that if a barcoded ancestor was used in the competition experiments, a large number of reads (up to 90% of reads for the initial time point) would be spent sequencing the ancestral clone, leading to a waste of sequencing capacity when estimating the frequency of the 4,800 evolved clones in the pool. We attempted to use a barcode-less clone for the ancestor, but found that the PCR reactions failed when the barcode sequence was present in such a small proportion of the population. Therefore, we developed a barcoded ancestral strain with a restriction site at the barcode locus to serve as the reference strain, allowing us to remove the amplicons derived from the reference strain by restriction enzyme digestion after the PCR step, saving us a significant amount of sequencing cost.

We used the following primers in constructing the modified ancestor:

RE-SbfI-F	atcg cctgcagg aaacgaagataaatcatgtc
RE-ApaLI-2R	atcg gtgcac ctgtcaacactgttccaact
RE-ApaLI-2F	atcg gtgcac ataacttcgtataatgtatg
RE-XhoI-R	atcg ctgcag tcatgtaattagttatgtca

We used the plasmid pBAR3 ([Levy et al., 2015](#)) as a template to generate two separate PCR products from the above primers, which were then digested and ligated together to form the final construct for transformation. Primers RE-SbfI-F and RE-ApaLI-2R were used to generate amplicon A, while primers RE-ApaLI-2F and RE-XhoI-R generated amplicon B. Amplicon A was then digested with SbfI and ApaLI, amplicon B was digested with ApaLI and XhoI and the pBAR3 plasmid was digested with SbfI and XhoI. These three digestion products were mixed and ligated simultaneously to generate complete plasmids containing the ApaLI site in the barcode region. The ligation product was transformed directly into SHA185 ([Levy et al., 2015](#)) to generate the modified ancestral clone. The presence of the ApaLI site in the barcode locus was verified through amplification of the barcode locus of these transformants, digestion of the resulting product by ApaLI and gel electrophoresis. A number of validated modified ancestral clones were screened for ancestral fitness using the fluorescent pairwise-competition fitness assays (described in [Levy et al. \(2015\)](#)), and the clone with the fitness closest to ancestral was selected for use in the sequencing based fitness measurement assays.

### Optimizing Sequencing Costs

We determined that we needed  $\approx 1$  million reads per time point for Illumina amplicon sequencing of the 4,800-pool barcodes in order to accurately estimate the frequency of the clones. For the initial time point, we expect  $\approx 200$  reads per clone at this read depth. The frequencies of clones with large fitness effects (either positive or negative) are expected to change the most between samples (taken every 8 generations). A  $\approx 20\%$  fitness effect (the largest fitness effect observed in [Levy et al. \(2015\)](#)) would result in a 4-fold change in frequency over 8 generations, resulting in the subsequent sample having  $\approx 800$  reads for the clone if it is adaptive and  $\approx 50$  reads for the clone if it is deleterious. These read depths are sufficient to have a small amount of sampling noise during the amplicon sequencing process ([Levy et al. \(2015\)](#) Supplementary Methods section 5), and thus  $\approx 1$  million reads per sample are adequate for our purposes. After allowing for variation in read depth when multiplexing samples (one sample per time point), the presence of a small amount of reads from the ancestral reference strain (due to low levels of undigested PCR products) and the use of a 25% Phi-X library spike-in to properly calibrate the Illumina machines, we ended up pooling 9 assays worth of samples ( $\approx 40$  samples) per lane of Illumina HiSeq 2000. As each lane results in  $\approx 200$  million reads, this gave us  $\approx 4$  million raw reads per sample. After removing reads from the reference strain and PCR duplicates, we had 1-3 million reads per sample for our estimation.

As we sequenced 5 time points per fitness assay replicate, this protocol costs  $\approx \$0.06$  USD to measure the fitness of a single clone per replicate. It takes a single person about one month to conduct both the fitness measurement assays and library preparation for amplicon sequencing, showing that this is truly a fast, accurate, and cost-effective way to estimate the fitness of thousands of clones in parallel.

### Pooling the 4,800 Clones Sampled at Generation 88

The 4,800 sampled clones from generation 88 of the two evolution replicates of [Levy et al. \(2015\)](#) were stored in glycerol stocks in 50 96-well plates. As it was impractical to pool all of these clones together at once, we constructed the pool in batches of 192 clones (2 plates). Each clone was grown from freezer stock in 800  $\mu\text{l}$  of M3 medium (the medium used in the evolution experiments) in 96 well plate format at 30°C for 2 days so that all lineages reached saturation. 400  $\mu\text{l}$  40% glycerol were added to each well and mixed, after which 400  $\mu\text{l}$  mixture from each well were pooled into a single vessel. Thus, for every two plates (192 total clones) we had an 80mL pool stored in two 50mL tubes at  $-80^\circ\text{C}$ . This procedure was repeated for pairs of plates over the course of a few weeks until we had 25 frozen pools, each of which represented two plates worth of clones. As we used a multi-pronged pinner to take clones from frozen stock and pin them into 96-well plates, a small percent of clones were not successfully recovered from frozen stock and therefore not included in the pool. The 25 frozen pools were then thawed simultaneously at room temperature and mixed into a single vessel. This vessel thus contained cells from  $\approx 4,800$  clones (excluding those that were not recovered from frozen stock). We dispensed 1 ml aliquots into 1.5mL eppendorf tubes, which were stored at  $-80^\circ\text{C}$ .

We found that the clones have a wide range of frequencies in the pool, spanning nearly 3 orders of magnitude. To test whether this wide frequency range had a significant effect on fitness, we generated another pool of 500 of these clones where all clones were grown and pooled simultaneously, instead of in batches, and the fitness assay was begun without any freeze-thaw cycles to minimize the number of generations of pooled growth before the beginning of the fitness assay. Our fitness measurement results are highly consistent with the results of the 4,800 clone pool ([Figure 2](#)), suggesting that the wide range of initial frequencies, freeze-thaw effects nor the presence of additional generations of growth in the pool substantially change our fitness estimates.

### Conducting the Fitness Measurement Assay

The fitness measurement assay was designed to assay the fitness of a large number of adapted clones in bulk against a reference clone. We conducted the fitness measurement assay on the pool of 4,800 clones in four batches with slightly different protocols.

To conduct the fitness assays, we first streaked the modified ancestral clone from frozen stock onto M3 agar plates. We selected a single colony and inoculated it into 3mL fresh M3 media and grew it for 2 days so that it reached saturation. 400  $\mu\text{l}$  of cell culture were then inoculated into 100mL M3 medium (the medium used in the evolution experiments of [Levy et al. \(2015\)](#)) in 500mL DeLong flasks (Bellco # 2510-00500). We also thawed out 1mL of the 4,800 clone pool, spun it down, removed supernatant, re-suspended the cells in M3 medium (to remove glycerol) and then inoculated the entire volume into a separate flask of 100mL M3 medium. After 2 days of growth at 30°C and 223 RPM in a shaking incubator, the cultures were saturated, and we mixed the ancestral culture with the pool in a 1:9 ratio accounting for variation in particle counts between the two cultures (Beckman Coulter) resulting in  $\approx 100\text{mL}$  of mixture. 400  $\mu\text{l}$  of this mixture were then used to inoculate 3 replicate fitness assay cultures. The replicate fitness assay cultures were grown under conditions identical to the initial evolution conditions ([Levy et al., 2015](#)) for a total of 4 growth cycles or 32 generations with 1:250 dilutions at every transfer. The remainder of the 100mL culture after the initial mixture and after each transfer was aliquoted in two 50mL conical tubes, spun down at 3000 rpm for 5 min, re-suspended in 6mL sorbitol solution (0.9M sorbitol, 0.1M Tris-HCL [pH 7.5], 0.1M EDTA [pH 8.0]) and frozen at  $-20^\circ\text{C}$  ( $-80^\circ\text{C}$  is also acceptable). This procedure was done for three different batches of assays (batches # 1, # 3 and # 4). The 500 clone pool measurements followed a similar protocol except we did not conduct the recovery growth from the freezer stock, as the 500 clone pool fitness assays were conducted without freezing the population (so one two-day growth cycle between the initial pooling and the mixing of the pool with the ancestor to begin the fitness assay).

For the batch # 2 containing two replicates (the third replicate did not generate sufficient sequencing data for analysis), after the initial 2 day growth of the separate ancestral and pool cultures in 100mL M3 media performed as for the first batch of experiments, we transferred  $5 \cdot 10^7$  cells from each culture into 100mL of fresh M3 medium and grew them separately for 2 days before mixing and

beginning the assay as before. This second 48 hr growth was done to accustom the cells to the medium and to minimize the freezer effects before beginning the assay. The number of cells/mL was determined using a Coulter particle counter to transfer  $5 \cdot 10^7$  cells for each transfer, rather than the 400  $\mu$ L transfers done by Levy et al. (2015). This was done to ensure a more consistent dilution regime, and in practice worked out to nearly the same regime as the evolution experiments as we transferred  $\approx 400 \mu$ L per cycle under these conditions.

In effect, while all three batches were tracked for 32 generations of growth, we used the data from generations 8-32 for fitness estimation in batches 1 and 3, and data from generations 0-24 in batch 2.

### DNA Extractions from Each Sample

For each sample (representing one time-point in one replicate), we conducted DNA extractions as follows (starting from 50mL of cells spun down, then re-suspended and frozen in 5mL of sorbitol solution: 0.9M sorbitol, 0.1M Tris-HCL [pH 7.5], 0.1M EDTA [pH 8.0]). We thaw the frozen samples at room temperature, resuspend the cells by vortex and transfer 750 $\mu$ L of cells to a 2mL screw cap tube. The cells are then collected by high speed centrifugation, the supernatant is removed and the cells are washed in 500  $\mu$ L sterile H<sub>2</sub>O. The water is again removed by centrifugation. We then add 200  $\mu$ L Triton SDS buffer (2% (v/v) Triton X-100, 1% (w/v) SDS, 100mM NaCl and 1mM Na<sub>2</sub> EDTA) to the cells, along with 200  $\mu$ L 25:24:1 phenol: chloroform: isoamyl alcohol and  $\approx 200 \mu$ L 0.1mm glass beads. This mixture is vortexed at high speed for 15 min. We then add 200  $\mu$ L (pH8.0) TE buffer to the tubes in a fume hood, then spin the tubes for 2 min in a microcentrifuge at high speed to collect the cellular debris. The aqueous layer is transferred to a 2mL yellow phase lock tube (5 PRIME # 2302830), which is then spun for 5 min at high speed in a microcentrifuge. The supernatant from the phase lock tube is transferred to a clean 2mL eppendorf tube, along with 1mL cold 100% ethanol. This is mixed by inversion, which should visibly precipitate the DNA. The DNA is collected by centrifugation for 2 min at high speed, after which the supernatant is discarded. The DNA pellet is resuspended in 400  $\mu$ L TE buffer, to which we add 50  $\mu$ L 10mg/mL RNase A and incubate for 15 min at 37°C. We add 10  $\mu$ L 4M ammonium acetate plus 1mL 100% ethanol to the mixture and mix by inversion. The DNA is collected again by centrifugation for 2 min at high speed, after which we remove the supernatant and let it air dry for 2 min before finally re-suspending the pellet in 150  $\mu$ L EB buffer (10mM Tris-Cl [pH 8.5]). We dilute this re-suspended DNA to 75ng/ $\mu$ L in EB for use in the PCR reactions (lower yields are acceptable as long as the concentration is at least  $\approx 40$ ng/ $\mu$ L).

### PCR Amplification of the Barcode Locus

We used a two-step PCR protocol to amplify the barcodes from the DNA that is very similar to the protocol used in Levy et al. (2015).

We use barcoded primers for the first PCR cycle. Different combinations of forward and reverse primers are used for each sample so that we can multiplex many samples together in a single HiSeq lane. The “N” positions in these primers are random nucleotides used to uniquely index each amplicon product to remove PCR duplicates from downstream analysis. All of these primers are HPLC purified to ensure that they are the correct length.

Forward primers

FP1	ACACTCTTCCCTACACGACGCTCTCCGATCT NNNNNNNN CGATGTTT AATATGGACTAAAGGAGGCTTTT
FP2	ACACTCTTCCCTACACGACGCTCTCCGATCT NNNNNNNN ACAGTGTT AATATGGACTAAAGGAGGCTTTT
FP3	ACACTCTTCCCTACACGACGCTCTCCGATCT NNNNNNNN TGACCATT AATATGGACTAAAGGAGGCTTTT
FP4	ACACTCTTCCCTACACGACGCTCTCCGATCT NNNNNNNN GCCAATTT AATATGGACTAAAGGAGGCTTTT
FP5	ACACTCTTCCCTACACGACGCTCTCCGATCT NNNNNNNN ATCACGTT AATATGGACTAAAGGAGGCTTTT
FP6	ACACTCTTCCCTACACGACGCTCTCCGATCT NNNNNNNN CAGATCTT AATATGGACTAAAGGAGGCTTTT
FP7	ACACTCTTCCCTACACGACGCTCTCCGATCT NNNNNNNN GGCTACTT AATATGGACTAAAGGAGGCTTTT
FP8	ACACTCTTCCCTACACGACGCTCTCCGATCT NNNNNNNN TAGCTTTT AATATGGACTAAAGGAGGCTTTT

Reverse primers

RP1	CTCGGCATTCTGCTGAACCGCTCTCCGATCT NNNNNNNN TATATACGC TCGAATTCAAGCTTAGATCTGATA
RP2	CTCGGCATTCTGCTGAACCGCTCTCCGATCT NNNNNNNN CGCTCTATC TCGAATTCAAGCTTAGATCTGATA
RP3	CTCGGCATTCTGCTGAACCGCTCTCCGATCT NNNNNNNN GAGACGTCT TCGAATTCAAGCTTAGATCTGATA
RP4	CTCGGCATTCTGCTGAACCGCTCTCCGATCT NNNNNNNN AACTGCGT TCGAATTCAAGCTTAGATCTGATA
RP5	CTCGGCATTCTGCTGAACCGCTCTCCGATCT NNNNNNNN ACTAGCAGA TCGAATTCAAGCTTAGATCTGATA
RP6	CTCGGCATTCTGCTGAACCGCTCTCCGATCT NNNNNNNN TGAGCTAGC TCGAATTCAAGCTTAGATCTGATA
RP7	CTCGGCATTCTGCTGAACCGCTCTCCGATCT NNNNNNNN CTGCTACTC TCGAATTCAAGCTTAGATCTGATA
RP8	CTCGGCATTCTGCTGAACCGCTCTCCGATCT NNNNNNNN GCGTACGCA TCGAATTCAAGCTTAGATCTGATA

For the first cycle, for each sample, we performed 12 PCR reactions.

Master Mix:

- 325  $\mu$ l OneTaq 2x master mix (NEB # M0482L)
- 13  $\mu$ l 10uM FP
- 13  $\mu$ l 10uM RP
- 156  $\mu$ l sample DNA (diluted to 75ng/ $\mu$ L or the entire DNA sample if between 40 - 75ng/ $\mu$ L DNA)
- 143  $\mu$ l dH<sub>2</sub>O
- 650  $\mu$ l total

50  $\mu$ l of master mix is aliquoted into 12 wells of a 96 well plate and the following PCR reaction is run on a thermocycler:

1. 94°C 10 min
2. 94°C 3 min
3. 55°C 1 min
4. 68°C 1 min
5. Repeat steps 2-4 for a total of 3 cycles
6. 68°C 1 min
7. Hold at 4°C

We then add 250  $\mu$ l of P1 buffer from QIAquick PCR purification kits (QIAGEN # 28106) to each PCR reaction and then perform PCR cleanups following the standard QIAGEN protocol in two columns (6 PCR reactions pooled into each column). This results in 50  $\mu$ l eluate of purified PCR product in two tubes for each sample.

For the second step of PCR, we use the following HPLC purified primers (where x is a phosphothioate group)

PE2 - xAATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCxT (read1)

PE1 - xCAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCxT (read2)

Master mix:

- 175  $\mu$ l PrimestarMAX 2x master mix (Clontech # R045B)
- 7  $\mu$ l 10  $\mu$ M PE1
- 7  $\mu$ l 10  $\mu$ M PE2
- 90  $\mu$ l purified PCR product
- 71  $\mu$ l dH<sub>2</sub>O
- 350  $\mu$ l total

50  $\mu$ l of this master mix is added to each of 6 wells in a 96 well PCR plate, and the following reaction is run:

1. 98°C 2 min
2. 98°C 10 s
3. 69°C 15 s
4. 72°C 15 s
5. Repeat steps 2-4 for a total of 24 cycles
6. 72°C 1 min
7. Hold at 4°C

250  $\mu$ l of Buffer P1 from the Quiagen kit is again added to each of these PCR wells, and all 6 wells are used in a single PCR purification protocol to generate a single tube with 50  $\mu$ l eluent with purified PCR product from the sample.

### Removal of the Reference Strain Amplicons Using Restriction Digestion and Size Selection

We conducted the ApaLI digest of the reference strain reads as follows. We added 60  $\mu$ l H<sub>2</sub>O, 10  $\mu$ l 10XCutSmart buffer (NEB # B7204S), 5  $\mu$ l ApaLI enzyme (NEB # R0507S) and 25  $\mu$ l of the purified PCR product to a single tube and digested for 2 hr at 37°C. After digestion, we did a standard PCR purification using the QIAGEN QIAquick PCR purification kit (QIAGEN # 28106) on the 100  $\mu$ l of digestion mixture and eluted in 30  $\mu$ l of buffer EB. After digestion, we conducted size selection using E-Gels (ThermoFisher # G661002) with 25  $\mu$ l of the purified digestion product and selected the band at  $\approx$ 350 bp for sequencing.

### Multiplexing and Amplicon Sequencing

We used Qubit HS kits (ThermoFisher # Q-33120) to quantify the concentration of our size-selected product for each sample and mixed them in equimolar ratios into a single sample for high-throughput Illumina sequencing. Our samples were submitted to the Stanford PAN facility (<http://pan.stanford.edu>) for Bioanalyzer analysis and then sequenced either with NGX Bio [www.ngxbio.com](http://www.ngxbio.com) or at the Stanford Center for Genomics and Personalized Medicine (<http://scgpm.stanford.edu>) with 2x101 paired end sequencing technology on Illumina HiSeq 2000 machines. Samples were sequenced with 25% phi-X genomic library spike-in (provided by the sequencing facility) to avoid calibration problems due to amplicon sequencing.

### Initial Processing of the Amplicon Sequencing Data

Our initial processing of the sequencing data included de-multiplexing the sequencing data to separate reads from different samples, removing PCR duplicates, and determining the number of reads in each sample for each barcode. Complete source code can be found at <https://github.com/sunthedeep/BarcodeCounter>.

Briefly, the pipeline uses bowtie2 to identify the sample, pcr duplicate, and lineage tag barcode sequences from each read in the FASTQ file. After removing PCR duplicates from the data and demultiplexing the data by sample, we identify all unique sequences in each sample and their number of occurrences using a simple lookup table. We then map all of these unique sequences to the database of 500,000 barcode sequences identified by Levy et al. (2015) using NCBI blastn with parameters (“-outfmt 6 -word\_size 12 -evaluate 0.0001”) to count the number of reads mapping to each of the known 500,000 barcodes in each sample. We account for barcodes known to be in the database with nearly identical sequences by considering such barcode clusters as a single lineage, and provide scripts to identify previously undetected barcode clusters from the sample data. These barcode counts provide the input for our fitness estimation procedure described below.

### Whole-Genome Sequencing

#### DNA Extraction, Library Construction, and Whole-Genome Sequencing

Clones selected for sequencing were streaked onto either M3 or YPD agar plates from freezer stocks for single colonies. One single colony for each clone was inoculated into either 1mL M3 or YPD (in a 96 deep-well plate) and grown overnight at 30°C without shaking. These cultures were used to perform DNA extractions using either the BioBasic 96 yeast genomic DNA extraction kit (BioBasic # BS8357) or the Zymo YeaStar Genomic DNA kit (Zymo # D2002). Libraries were constructed using Nextera technology with the protocol of Kryazhimskiy et al. (2014). We multiplexed up to 96 libraries per Illumina HiSeq 2000 lane; samples were sequenced at the Stanford Center for Genomics and Personalized Medicine with 2x101 paired end sequencing technology. Libraries that generated less than 5x average genome-wide coverage were removed from further analysis. Some lineages (defined by unique barcode IDs) were sequenced multiple times, either due to low coverage in one library or due to sequencing multiple independent clones containing the same barcode ID. Variants called from all libraries with the same barcode ID, regardless of origin, were combined together. Importantly, please note that while the libraries were mapped to a non-reference genome which includes the barcode locus sequence, all variants reported in this manuscript both in the main text and the supplemental files have been lifted over to the coordinate system of the *Saccharomyces* Genome Database (SGD; <http://www.yeastgenome.org>) R64 *Saccharomyces cerevisiae* reference genome for convenience.

#### FASTQ Processing, GATK-Based Variant Calling, and Filtering

For each sample, we received two fastq files, one for each read of the paired end sequencing (“forward.fastq” and “reverse.fastq”). We trimmed the first 15 bases and the last 3 bases of each read as well as any adaptor sequences using TrimGalore (version 0.3.7 Available at: [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)).

```
perl trim_galore -a CTGTCTCTTATACACATCT -a2 CTGTCTCTTATACACATCT -length 50 --clip_R1 15 --clip_R2 15 --three_prime_clip_R1 3 --three_prime_clip_R2 3 --paired -o OUTPUTDIR forward.fastq reverse.fastq
```

Reads were mapped using Novoalign (version 3.02.02, Novocraft Technologies) to a modified version of the *sacCer3 S288C S. cerevisiae* reference genome that includes the DNA barcode locus (Levy et al., 2015) in the sequence.

```
novoalign -d referenceGenome.fasta -f forward.trimmed.fastq reverse.trimmed.fastq -l 75 -H22 -o SAM READGROUPINFO -r Random library.novoalign.sam
```

The mapped reads were then sorted using PicardTools version 1.105(1632) (Broad Institute, <http://broadinstitute.github.io/picard>)  
java -Xmx2g -jar SortSam.jar INPUT=library.novoalign.sam OUTPUT=library.novoalign.bam SORT\_ORDER=coordinate

We used PicardTools again to remove PCR duplicates

```
java -Xmx2g -jar MarkDuplicates.jar ASSUME_SORTED=true REMOVE_DUPLICATES = true INPUT=library.novoalign.bam OUTPUT=library.novoalign.dedup.bam
```

After building an index for the bam file with PicardTools, we then got global coverage metrics with GATK version 3.2.2 (McKenna et al. 2010) DepthOfCoverage function and per base pair coverage statistics using Bedtools v2.17.0 Quinlan et al. 2010 genomecov. We plotted the per base pair coverage statistics to identify whole chromosome aneuploidy events.

We genotyped the libraries using GATKs Unified Genotyper.

```
java -jar -Xmx2g GenomeAnalysisTK -T UnifiedGenotyper -R referenceGenome -l library.novoalign.dedup.bam -ploidy 2 -genotype_likelihoods_model BOTH -stand_call_conf30 -stand_emit_conf 10 -o library.gatk.vcf
```

We initially filtered variants from GATK as follows:

```
java -Xmx2g -jar GenomeAnalysisTK.jar -T VariantFiltration -R referenceGenome --variantlibrary.gatk.vcf --out library.gatk.filtered.vcf --filterExpression "QD < 10.0 || FS >20.0 || MQ < 50.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || AN > 10 || AF<0.25" --filterName "my_filter"
```

We also manually filtered all variants called by GATK by remapping them with the CLC genomics workbench and manually validating variants using the resulting read pileups. This procedure eliminated 11% of our GATK variant calls. We then tested the validity of 57 of these filtered variants using Sanger sequencing, and identified no false positive calls. We combined all of the filtered variants across all of the libraries into a single file ("allLibraries.gatk.filtered.vcf") using GATK after lifting over the variant coordinates to the standard UCSC sacCer3 reference genome. We then conducted additional variant filtering using custom scripts, where we removed any mitochondrial variants, variants not passing the GATK filter and variants annotated in reference genome as being in repetitive elements (telomeres, centromeres, replication origins, transposable elements containing "Ty," "delta," "sigma" or "tau" in their name) or low complexity regions defined by the Tandem Repeat Finder (Benson, 1999) with the recommended parameters (2 7 7 80 10 50 500 ngs). We also removed all variants with less than 3 reads of support for the derived allele. Heterozygous calls by GATK were validated by first testing whether they had at least 3 reads of support for both ancestral and derived alleles, and passed a binomial filter with  $p > 5\%$  for deviation from an equal proportion of ancestral and derived reads. Variants that failed either of these filters were reclassified as homozygous. Heterozygous calls that did pass were then checked to see if they resided in homopolymer repeat regions or in sites with multiple derived alleles across the entire dataset. Such variants were removed from the dataset as likely mapping errors.

As we found that mutations in the nutrient sensing pathway were highly adaptive, we searched the raw variant calls of clones with  $s > 5\%$  but no nutrient sensing pathway mutations for filtered variants in this pathway and added them back into our mutation list (the mutations reported in the main text include these variants). This was done for a total of 3 clones (one *IRA1*, one *IRA2* and one *CYR1*).

### Copy-Number Variant Detection

We tested for the presence of copy-number variants using a number of software packages, including CNVnator and SVDetect, along with specific manual surveys of the coverage density around the *HXT6/7* locus as amplifications of this locus have been shown to be adaptive in previous chemostat laboratory evolution experiments. However, we were unable to detect any high-confidence copy-number events either at this locus or genome-wide.

### Structural Variant Detection with CLC-Bio

We systematically looked for the existence of structural variation in our sequenced clones, i.e., for the presence of insertions and/or deletions larger than the maximum of 5-10 bp typically detected by our GATK-based variant calling pipeline, as well as chromosomal inversions and translocations. We performed a workflow, described below, utilizing CLC Genomics Workbench version 8.5 (QIAGEN Aarhus A/S; [www.clcbio.com](http://www.clcbio.com); API version:850; Build number:20150904114350; Build date:1509041143; Build rev:131279. Platform:Mac OS X 10.10.5; Architecture:x86\_64 (64 bit); Processor cores:24; Java version:1.8.0\_60 (Oracle Corporation)). Note that we will call the program "CLC Workbench" for brevity. First we imported the Illumina paired end fastq.gz files for each clone into CLC Workbench, using the parameters "paired reads," "remove failed reads," "paired-end (forward-reverse)," minimum distance 25, maximum distance 1000, Illumina pipeline 1.8 and later quality scores.

We then mapped the reads to the unmodified *S. cerevisiae* (strain S288C) reference genome (downloaded from the *Saccharomyces* Genome Database (SGD; [www.yeastgenome.org](http://www.yeastgenome.org)) R64-1-1 and then imported into CLC Workbench). We did not use any masking during the mapping and used the following mapping parameters: mismatch cost 2, lineage gap cost, insertion and deletion costs 3, length fraction 0.5, similarity fraction 0.5, auto-detect paired distances, map randomly for non-specific matches.

Reads were then trimmed by using the "Trim Sequences" function; trimming was done based on quality scores (limit 0.05); ambiguous nucleotides (maximum of 2) were also trimmed. Reads below 15 nucleotides in length were discarded. Any Nextera adaptor sequences were trimmed from reads using the following sequence and parameters for trimming: sequence for adaptor trimming CTGTCTTATACAC, strand "plus," remove adaptor, mismatch cost 2, gap cost 3, allow internal matches with minimum score 4, allow end matches with minimum score at end 1.

We then ran the "InDels and Structural Variants" function, using these mapped and trimmed reads, with the parameters "p value threshold 0.001" and "maximum number of mismatches 3," and saved the "breakpoints" output files in tab-delimited formats. These variants were filtered to remove structural variants with less than 3 reads of support, present in more than 3 strains or closer than 300bp from the ends of each chromosome. The variants were annotated with gene annotations (file SGD\_features.tab) from the *Saccharomyces* Genome Database ([www.yeastgenome.org](http://www.yeastgenome.org)).

After structural variant calling was completed, we filtered out structural variants that occurred in previously known repetitive elements annotated in the SGD database (telomeres, centromeres, replication origins, transposable elements containing “Ty,” “delta,” “sigma” or “tau” in their name) as before.

## Determination of Mating Type and Ploidy

### Mating Type Assays

Mating type testing was conducted for 960 clones from replicate E2 and 192 clones from replicate E1. Standard Nat+ *URA3*- tester strains of both MAT $\alpha$  and MAT $a$  mating types were grown as lawns on YPD agar plates, while the clones with unknown mating type (Nat-, *URA3*+) were arrayed and grown on independent YPD agar plates. Replica plating was used to transfer the clones with unknown mating type onto Nat+ Ura- SC-agar plates along with one of the tester strains. The presence of a colony on this plate was used to determine successful mating.

### Propidium Iodide and Flow Cytometry

Ploidy was initially tested using a simplified propidium iodide staining protocol designed for high throughput analysis, inspired by Cousin et al. (2009). Clones were grown to saturation in YPD liquid media in 96 well plate format. 200 ml of saturated culture was transferred to 96 well filter plates (Pall Life Sciences # 8039) and spun down to remove the spent media. These spun down cells were resuspended in 200  $\mu$ l 70% ethanol in the filter plates and allowed to fix for at least 1 hr at room temp. Plates were then centrifuged again to remove the ethanol. Cells were resuspended in 50  $\mu$ l RNase A buffer (1mg/mL RNase A in PBS) and incubated at 37°C for at least 6 hr (at most 18 hr). Treated cells were diluted 1:100 into 200 L of propidium iodide staining solution (50  $\mu$ g/ml PI, 50  $\mu$ M sodium citrate) and analyzed along with standards of known ploidy using the BD LSR II with an HTS attachment at the Stanford Shared FACS Facility (NIH grant # S10RR027431-01 for UV LSR II). We note that the filter plates can be re-used for ploidy analysis by thoroughly washing them with distilled water using a multichannel pipette.

### High Throughput Benomyl Assay

A simpler high throughput ploidy test was developed using the drug benomyl. Clones were grown from frozen stock in 1mL liquid YPD in 96 well plates until saturation at 30°C without shaking. The saturated cultures were mixed by multichannel pipette, pinned onto YPD+20  $\mu$ g/mL benomyl (in DMSO) and YPD+DMSO (control) rectangular agar plates using a multi-pronged pinner, grown at 25°C for 48 hr, and then imaged. Under these conditions, diploid growth is strongly inhibited by benomyl but haploid growth is less affected.

## Construction of Gene Deletions in the Ras/PKA Pathway

Gene deletions were constructed using standard yeast transformation methods to replace the gene of interest with a selectable marker cassette. *IRA1*, *IRA2*, *GPB1*, *GPB2*, *PDE1*, *PDE2* and the pseudogene control YFR059C were individually replaced with a selectable NatMX (nourseothricin) resistance marker in neutral barcoded yeast strains. For each target gene, the resistance marker was amplified from the pBAR1 plasmid (Levy et al., 2015) with primers flanked by 45 bp of sequence adjacent to each end of the appropriate yeast gene. Transformations were performed to delete the gene of interest using the lithium acetate based protocol of Gietz and Woods (2002). Each transformant was verified with gene-specific PCR reactions spanning both the 5' and 3' insertion breakpoints. We assayed the fitness of each deletion using pairwise competition assays described in Levy et al. (2015) and in the main text methods.

Primers for gene deletions:

```
IRA1 5' CTTCAGCATATAACATACAACAAGATTAAGGCTCTTTCTAAAATGTGGAGGCCAGAAATACCCTCC
IRA1 3' AAGGAAAAACGTATATAATCACTGCAACTCTAATTTAAAATTATCGACACTGGATGGCGGC
IRA2 5' TATCAACTAAACTGTATACATTATCTTTCTTCAGGGAGAAGCATGTGGAGGCCAGAAATACCCTCC
IRA2 3' AGATAGATATTGATATTTCTTTTCATTAGTTTATGTAACACCTCTATCGACACTGGATGGCGGC
GPB1 5' CGGCTACTTTAAGGCTTTCCGTACCAATTCTTCTACATAAGAATGTGGAGGCCAGAAATACCCTCC
GPB1 3' AATTTTCTCGTTTTCCTTTAGTCACTCTTGTACATAAAGGATTATTCGACACTGGATGGCGGC
GPB2 5' GATTCATTGGCAGGTCCATTGTCGCATTACTAAATCATAGGCATGTGGAGGCCAGAAATACCCTCC
GPB2 3' CTAACAAGTTTACAAAGTGAAGCATTGAAACTGCCTTTTTATCGACACTGGATGGCGGC
PDE1 5' GGTTCCTTCTTTCATCCCCTTTTTACCAATATTCCTTTTTATGTGGAGGCCAGAAATACCCTCC
PDE1 3' TAATGAAAAGAAGTTTCATTAGTACTACTAGTATTTTGCTTGCTTCGACACTGGATGGCGGC
PDE2 5' GAGATCACTACTTAATTGAAGAAAACATAACCTATTGATATGTGGAGGCCAGAAATACCCTCC
PDE2 3' ATGTTTATACAATGAATGGTACAAGAAATTTTGATATCTTGCTATCGACACTGGATGGCGGC
YHR095W 5' CCATCAAATGTGCGCAGCAGCTCATGTTTACGTTTGCTGTCTTCTGTGGAGGCCAGAAATACCCTCC
YHR095W 3' AATAAGCCCTAGAAACCTTACACCCTAATTTGCACAAGAAAACCTATCGACACTGGATGGCGGC
```

## QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters including the exact values of *n*, precision measures (mean  $\pm$  SEM) and statistical significance for various statistical tests are reported in the Main Text, Figures and the Figure Legends.

### Introduction to Fitness Estimation Methodology

The fitness of a barcoded lineage relative to the rest of the population determines how quickly it grows. If the number of cells in a lineage is large at the bottleneck, then during the  $T=8$  generations from cycle  $i$  to cycle  $i+1$  the bottleneck population,  $n$ , grows close to deterministically:

$$n_{i+1} \approx n_i e^{(s-\mu_i)T} \quad (1)$$

with  $\mu_i$  the mean fitness of the population at time  $i$ . The time-dependent mean fitness cannot be measured directly, but the size of the total barcoded population that is neutral with respect to the ancestor,  $\rho_i$ , gives  $\mu_i$  from  $\rho_{i+1}/\rho_i$  as  $s$  is the fitness of the barcoded lineage relative to these neutral lineages.

The sequencing measurements give estimates of the relative sizes of a barcoded lineage from the numbers of reads,  $r_i$ , of the barcode at successive time points as a fraction of the total reads,  $R_i$ . Comparing with the number of reads of the neutral barcodes,  $\rho_i$ , the fitness over cycle  $i$  is estimated by

$$\begin{aligned} \hat{s}_i &= \frac{1}{T} [\ln(r_{i+1}/R_{i+1}) - \ln(r_i/R_i)] + \mu_i \\ &= \frac{1}{T} [\ln(r_{i+1}/r_i) - \ln(\rho_{i+1}/\rho_i)] \end{aligned} \quad (2)$$

However, there are several sources of deviations of such estimates from the actual fitness. The experiments themselves contribute biological stochasticity in the growth and division of cells, sampling during the dilution at the end of each cycle, and subtle variability in conditions. The measurement process contributes counting noise from sequencing as well as potential variabilities and biases in DNA extraction and PCR amplification.

The biological noise, dilution sampling, and sequencing counting noise should all have variance proportional to the mean numbers of cells and/or reads. We find that for typically sized barcode lineages ( $\sim 100$  reads), deviations from deterministic trajectories scale as the square root of the number of reads, i.e.

$$\text{Var}(r_i) \approx \kappa_i \langle r_i \rangle \quad (3)$$

where  $r_i$  is the number of reads at time  $i$ ,  $\langle r_i \rangle$  is the expected number of reads, and  $\kappa_i$  is a noise parameter inferred from the data which depends on the cycle, the replicate, and the batch. Furthermore, we show that for the collection of neutral lineages, the distributions of changes in read numbers from one cycle to the next are close to normal.

For large lineages ( $>10^3$  reads), however, the data exhibit larger than expected variations which do not decrease with numbers of reads. The sources of these variations are currently unknown. They set a limit of  $\geq 1\%$  per generation on the resolution of our fitness assay.

We use the data to crudely fit a multiplicative noise parameter  $\alpha_i$  at each cycle in addition to the normal variance. For the fitness inferred over one cycle,

$$\hat{s}_i = \langle s_i | r_i, r_{i+1} \rangle \quad (4)$$

the variance is then roughly of the form:

$$\text{Var}(s_i) = \frac{1}{T^2} \left( \frac{\kappa_i}{r_{i+1}} + \alpha_i^2 \right) \quad (5)$$

To infer fitnesses, we use a model assuming Gaussian additive noise at low frequency and multiplicative noise at high frequency to combine the results from across the cycles, replicates, and batches, weighted by the inverse variances.

In the next we further elucidate the fitness estimation process and break down the contributions to  $\kappa_i$ . We then carry out self-consistency checks and justify our noise model. Finally, we present the results of the fitness assay broken down by batch and replicate, and further discuss the hypothesis testing done in the main text.

### Noise Model

#### Read Stochasticity

From the dynamics of the numbers of cells in a lineage, we expect that the mean number of reads at time  $i+1$  will be

$$\langle r_{i+1} \rangle = \frac{R_{i+1}}{R_i} r_i e^{(s-\mu_i)T} \quad (6)$$

and thus dependent on the total numbers of reads,  $R_i$  and  $R_{i+1}$ .

The stochasticity in the population dynamics and the counting variations from the sequencing both give additive noise so that we expect

$$\text{Var}(r_{i+1}) \approx \kappa_i \langle r_{i+1} \rangle \quad (7)$$

where  $\kappa_i$  is a parameter fit from the data for many barcode families. For nearly neutral lineages, there is only a weak dependence of  $\kappa_i$  on  $s$ . Lineages with large  $s$  quickly reach a size where the additive model breaks down; we will use our multiplicative noise model to analyze their fluctuations.

The contributions to  $\kappa_i$  depend on the parameters of the particular measurement: the total number of barcoded cells at the bottleneck,  $\mathcal{N}_i^B$ , and the number of reads,  $R_i$ , both of which vary considerably. The average number of reads per barcoded cell, the *coverage ratio*

$$C_i \equiv \frac{R_i}{\mathcal{N}_i^B} \quad (8)$$

strongly affects the noise magnitude: when the coverage ratio is low, read noise dominates; when the coverage ratio is high, the biological noise dominates. The contributions to the noise parameter are

$$\kappa_i = \underbrace{1}_{\text{Read noise at } i+1} + \underbrace{C_{i+1}/C_i}_{\text{Read noise at } i} + \underbrace{C_{i+1}(\beta_i + 1)}_{\text{growth + dilution}} + \underbrace{\xi_i(1 + C_{i+1}/C_i)}_{\text{Extraction/PCR noise}} \quad (9)$$

The first term comes from Poisson read noise at time  $i+1$  and the second term from the Poisson read noise at time  $i$ , scaled from the coverage at  $i$  to the coverage at  $i+1$ . The third comes from the stochasticity in the growth of the cells. For a single cell at a bottleneck, the number of descendants at the end of the cycle averages  $2^T e^{sT}$  with a variance of  $\beta_i 2^{2T}$ . Almost all this variability is likely to come from the earliest stages of the cycle as when the number of descendants becomes large, the fluctuations are averaged over. After dilution the biological stochasticity contributes  $\beta_i$  per cycle to the variance. In addition, there is a factor of 1 that comes from the Poisson dilution at the end of each cycle. The last term  $\xi_i$  accounts for the unknown additive parts of the effects of DNA extraction and PCR amplification.

We assume that the variations are Gaussian in nature. This assumption was inspired by the additive nature of the noise sources, and describe the data well. The assumption breaks down when  $r_i$  is low or when  $C_{i+1}\beta_i$  is large, since the biological noise is likely to be non-Gaussian.

### Number of Mutants and Coverage Ratios

In order to understand the balance between read noise and biological noise, we need to know the coverage ratio,  $C_i$ , at each time point. We know the total number of reads,  $R_i$ , at each time point; however, we do not have a direct measurement of the total number of barcoded cells,  $\mathcal{N}_i^B$ , at the bottleneck of each cycle. Since the total population saturates at a size that is roughly independent of the admixture of mutants and ancestral types, after dilution the total bottleneck population,  $\mathcal{N}$ , is roughly constant.

The barcoded portion can be inferred by noting that two portions of the total population, the unbarcoded ancestral cells with population  $\mathcal{N}_i^U$ , and the barcoded types that are neutral relative to the ancestor at time  $i$ , have the same fitness. Given  $f_i^n$  the fraction of the barcoded cells without adaptive mutations, the ratio of the neutral population sizes,  $f_i^n \mathcal{N}_i^B / \mathcal{N}_i^U$  is thus constant.

If we know  $f_i^n \mathcal{N}_i^B / \mathcal{N}_i^U$  at one time point, and  $f_i^n$  at all other time points, we can solve for  $\mathcal{N}_i^B / \mathcal{N}_i^U$ . We can then use  $\mathcal{N}_i^B / \mathcal{N}_i^U$  and  $\mathcal{N}_i^U = \mathcal{N} - \mathcal{N}_i^B$  to approximate  $\mathcal{N}_i^B$  at each time point. Let  $f_i^n \mathcal{N}_i^B / \mathcal{N}_i^U = q$ . Then we have

$$\frac{f_i^n \mathcal{N}_i^B}{\mathcal{N} - \mathcal{N}_i^B} = q$$

$$\mathcal{N}_i^B = \frac{q}{q + f_i^n} \mathcal{N} \quad (10)$$

Initially, the fraction barcoded is formulated to be  $\mathcal{N}_0^B \cong 10\% \mathcal{N}$ . This gives allows us to calculate  $q$  ( $\approx 0.03$ , similar across batches). Using the sequencing reads, we can obtain  $f_i^n$  at each time point and hence calculate an estimate for  $\mathcal{N}_i^B$  at each time point.

Figure S1 shows this estimate of the barcoded fraction of the population. At late times, a considerable fraction of the population is barcoded, and a significant fraction of this barcoded population has adaptive mutations. The barcoded fraction increased rapidly as a consequence of its original diversity. Roughly 50% of the barcoded cells had fitness  $>6\%$  at the first time point, and 25% are diploid. The rest were nearly neutral haploids. By the end of the experiment,  $>90\%$  of barcoded cells were high fitness mutants. As barcoded fraction of the pool increased, the read depth remained nearly constant. The coverage ratios decreased as a function of time. They started at around 0.3-0.5, but fell to about 0.04-0.07 by the end of the experiment. We will see that means by late times, the errors in fitness estimation are dominated by the read noise.

The beneficial mutants showed significant transient behavior in the first growth dilution cycle. In batches 1, 3, and 4, the barcoded fraction (and therefore the beneficial mutants) did not increase appreciably in the first cycle. Batch 2 was grown for one growth/dilution cycle before time point 1 and does not display transient behavior in its first cycle.

To remove the effect of transient behavior on the fitness assay, we used the sequencing data from time points 2-5 for batches 1, 3, and 4, and time points 1-4 for batch 2. The trajectories of the barcoded fractions are very similar across batches for the time points chosen, and avoids the latest time point in batch 2 where the barcoded types have nearly taken over the population.

While  $\kappa_i$  increases with increasing coverage ratio, the variance in the fitness estimate decreases with increasing read depth. For large coverage ratios ( $C_{i+1} & C_i \gg 1$ ), the variance reaches the minimal value

$$\text{Var}(s_i) \approx \frac{1}{T^2} \frac{(1 + \beta_i)}{n_i} \quad (11)$$

where  $n_i$  is the number of cells in the barcode family at the bottleneck at time point  $i$ . In this regime the noise is dominated by biological fluctuations. This sets a noise floor for the measurement. For our measurements, only the first cycle (which was not included in the analysis) was near this regime.

### Inferences of $s$

In addition to the additive sources, there appears to be a roughly frequency independent component of the noise. The source of this noise is unknown. For simplicity, as it does not affect much the results, we parametrize this by a multiplicative Gaussian noise parameter  $\alpha_i$ , fit within each batch for every pair of time points. We find that  $\alpha_i \approx 0.1/\text{cycle}$ , largely independent of the cycle, replicate, and batch. Then the assumed variance of our estimator is

$$\text{Var}(s_i) = \frac{1}{T^2} \left( \frac{\kappa_i}{\langle r_{i+1} \rangle} + \alpha_i^2 \right) \quad (12)$$

The fitness estimation algorithm proceeds in the following manner:

1. Identify lineages which are neutral relative to the ancestor for each replicate and batch individually.
2. Use the *collection* of these neutral lineages to estimate  $\kappa_i$  and  $\mu_i$ .
3. Estimate  $\alpha_i$  for each batch and time point from lineages with a large number of reads.

We then carry out the follow steps for **each barcode separately**:

1. Use formulae for  $\hat{s}_i$  and  $\text{Var}(s_i)$  (Equations 2 and 12) to calculate fitness and error at each time point.
2. Average over time points, replicates, and batches, using inverse variance weighting by errors, to get an overall estimate of the fitness  $\bar{s}$  of that barcode.

We give a more detailed account in the next two sections.

### Checks on the Noise Model

We made a number of self-consistency checks to test the applicability of the simple additive noise model for lineages at low read depth. We analyzed the following quantities:

- Distributions of within-replicate variations
- Scaling with read numbers of between replicate variations
- Comparison of within replicate to between replicate variations

Our analysis suggests that there is good agreement between within-replicate variation and between-replicate variation for moderately sized ( $\sim 100$  reads) lineages. At late time points, the noise is dominated by the counting noise of sequencing. We show that this is due to the expansion of the barcoded lineages. We also discuss the frequency-independent deviations for large ( $\sim 1000$  reads) lineages, which limits the sensitivity of fitness assay to 10%/cycle (1.2%/generation).

### Estimating $\kappa$ within Replicate

By considering the dynamics of large groups of lineages with identical fitness together, we can test the noise model. The large set of lineages neutral relative to the ancestor ( $\sim 1500$ ) enables estimation of the noise parameter  $\kappa_i$ , with good enough statistics on the noise to test its normality. It also lets us infer the time-dependence of the mean fitness,  $\mu_i$ , which is needed to obtain the fitness of the other lineages. We assume that the neutral lineages are virtually identical in both fitness and the magnitude of their biological fluctuations.

The model assumes that the deviations of the read numbers  $r_{i+1} - \langle r_{i+1} \rangle$  are distributed as  $\mathcal{N}(0, \kappa_i \langle r_{i+1} \rangle)$  for  $r_{i+1}$  large enough. We define the normalized differences  $Z_i$  as follows:

$$Z_i = \frac{r_{i+1} - \langle r_{i+1} \rangle}{\sqrt{\langle r_{i+1} \rangle}} \quad (13)$$

Given a collection of lineages with identical phenotype, the  $Z_i$  are identically distributed as  $\mathcal{N}(0, \kappa_i)$ . The total frequency of a phenotype  $f_j$  can be used to estimate  $\mu_i$ . The distribution of scaled deviations can be used to find  $\kappa_i$  and test the noise model.

We carried out the following procedure using the  $\sim 1500$  lineages which were neutral in the experiments of [Levy et al. \(2015\)](#)

- Estimate mean fitness by  $\mu_i = \log(\bar{f}_{i+1}/\bar{f}_i)/T$ .
- Plot distribution of  $Z_i$ . Remove outliers (likely adaptive or outside regime of additive noise)
- Re-estimate  $\mu_i$ . Re-plot  $Z_i$ .
- Set  $\kappa_i = \text{Var}(Z_i)$ .

The first two rows of [Figure S2](#) shows the distributions of the scaled deviations for each pair of adjacent time points in batch 1, replicate 1. The top row shows the histograms of  $Z_i$  from the neutral haploids as a function of time. The same analysis was carried out on the 1600 diploid lineages: the results are shown in the second row and are very consistent with the haploid inferences.

The normal distribution predicted from theory is plotted in red over the empirical distribution histogram. The counting noise limit is plotted in black. The noise starts off larger than the read counting noise limit, but is dominated by counting noise at the end. This is in concordance with our observation that the coverage ratio decreases at late times, and suggests that the extraction and amplification parts of the noise,  $\xi_i$  — which would be expected also to scale with the read depth — is a small fraction of  $\kappa_i$ .

The model fits quite well over the range of 1-2 SDs. Only 10-20 lineages were removed from each plot as outliers; most were a few SDs from the mean, and clearly adaptive. The normal fit is worst at the earliest time points, when the cells are first experiencing the evolution condition, and the latest time point, where the number of reads is smaller. Our analysis also showed that the noise parameter inferred from lineages of different sizes did not vary significantly when different sized lineages were used to infer it (from 30 reads up to 150). The diploids behaved similarly to the haploids in all these aspects, including the number of outliers.

The values of  $\kappa_i$  tend to start around 10, and drop down to around 2 at late times. The values vary between replicates and experiments as can be expected by the different coverage ratios. Two replicates have very low coverage at one time point (batch 2 replicate 2 and batch 3 replicate 3), which decreases the quality of fitness inferences for those datasets. The analysis for the diploids found similar  $\kappa_i$  values. The diploid  $\kappa_i$  tended to be slightly higher than the haploid  $\kappa_i$ : by 5%–10% at early times, nearly identical at late times.

The fact that the fluctuations are dominated by counting noise at the end of the experiment suggest that  $\xi_i$  is small. If we set  $\xi_i$  to 0, and use the  $C_i$  estimated previously, we can calculate the biological noise parameters  $\beta_i$  using Equation 9. We get values  $\beta_i$  in the 12-17 range at early times and in the 5-10 range at late times.

Part of the difficulty estimating  $\beta_i$  comes from the fact that coverage ratios are low ( $\sim 0.1-0.2$ ). Therefore, errors in estimation of order 0.5 (from the mean fitness estimate, coverage ratio, and  $\xi_i$ ) propagate up to errors in  $\beta_i$  of order 2.5-5 at late times. More detailed analysis and measurements would need to be conducted to yield a more quantitative estimate for  $\beta_i$  and its uncertainty.

In previous experiments, the estimated values of  $\beta_i$  started off low (around 4), but reached values as high as 15 at later times when there were more mutants in the population. Since our experiments start off with a relatively high mutant fraction, our results are at least roughly consistent with previous work. Large values of  $\beta_i$  suggest that there is high variability when the populations are low: i.e., variations in viability (surviving stationary phase), lag phase (time delay to start growth after dilution), and in the first rounds of division.

### Replicate-Replicate Correlation

As an independent test of the consistency of the noise model, we examined the correlation between replicates in the same batch and compared to the inferred within-replicate noise parameters  $\kappa_i$ . Specifically, we looked at the sample SD of the log slope  $\sigma_i = \ln(r_{i+1}/R_{i+1}) - \ln(r_i/R_i)$ . The log slope was chosen since its variance can be shown to be

$$\text{Var}(\sigma_i) \approx \frac{\kappa_i}{r_{i+1}} \quad (14)$$

if our additive noise model holds with our definition of  $\kappa_i$ .

The final row of [Figure S2](#) shows the sample SD  $\delta\sigma_i$  of the log slopes plotted against the number of reads at the second time point of the cycle. The plots show the  $r_{i+1}^{-1/2}$  scaling as expected for a wide range of reads.

We can use the distribution of  $\delta\sigma_i$  to fit a  $\kappa$  parameter, and compare it to the expected value. The 3 curves with the scaling  $r_{i+1}^{-1/2}$  show 3 different fits. The within replicate  $\kappa$  from the variance of  $Z_i$  is shown in red. In blue is the inference  $\hat{\kappa} = E[\sqrt{r_{i+1}} \delta\sigma_i]^2$  (between replicate estimate). Black is the theoretical minimum value that the noise parameter could take if there was only read noise ( $\beta_i=0$ ).

For the first pair of time points,  $\kappa$  from within a replicate is larger by  $C_{i+1}/C_i$  compared to the between replicate  $\kappa$ . This is expected since the first measurement is common for all replicates in a single batch (see [STAR Methods](#)). By late times, both estimates of  $\kappa$  are close to the being pure read noise.

### Multiplicative Noise Regime

For each batch and time point, we roughly fit a frequency independent part of the noise by averaging  $\delta\sigma_i$  at high ( $\sim 10^3$ ) read number (green line). We then modify the noise parameter  $\kappa_i$  to be

$$\tilde{\kappa}_i = \kappa_i + \alpha_i^2 r_{i+1} \quad (15)$$

The multiplicative noise varies little across time points and experiments, typically  $\sim 10\%$ /cycle corresponding to uncertainties in estimates of  $s$  of  $\geq 1\%$  per generation. For typical values of  $k_i$ , the crossover between multiplicative noise and read noise occurs at  $\sim 10^3$  reads. The multiplicative noise constant increases with time in batch 2, which is dominated by mutants at late times.

## Fitness Assay

### Basic Procedure

We used the fitness assay outlined earlier to calculate fitnesses and error estimates for each lineage, between every pair of time points, replicates, and batches. For each replicate, we combined estimates across time points into an overall fitness estimate  $\bar{s}$  by an inverse variance weighted sum:

$$\bar{s} = \left( \sum_i \hat{s}_i / \text{Var}(s_i) \right) / \left( \sum_i 1 / \text{Var}(s_i) \right) \quad (16)$$

$$\text{Var}(\bar{s}) = \left( \sum_i 1 / \text{Var}(s_i) \right)^{-1} \quad (17)$$

This method gives the correct weighting for the max posterior estimate of the mean of a collection of Gaussian random variables with equal means and unequal variances.

We averaged over time points within a replicate to obtain the fitness values reported for each replicate. We averaged once more across replicates and batches to obtain the fitness values reported in the main text.

### Distribution of Fitness Effects

Figure S3 shows the distributions of fitness effects from all the replicates and batches. The two colors correspond to the haploids and the diploids respectively. As can be seen, almost all the lineages are either very close to neutral relative to the ancestor, or diploid. Both the neutral haploids and diploids tend to have low coverage at late times, which gives broad peaks (typically in the 2% range) compared to fitness differences (1.5%–3.5%). Due to systematic variation between batches, the diploid and main haploid peaks are well resolved only for some batches.

### Replicate-Replicate Fitness Correlation

To test the data against our noise model we examined the replicate-replicate fitness correlation. High fitness lineages tend to have lower errors due to higher read counts. Their errors are dominated by the multiplicative noise. We can see that the errors inferred via the noise model are very similar to the observed variation between replicates.

Figure S4 shows all of the replicate-replicate correlations. The scale of the inferred error bars is consistent with the scale of the differences in fitness between replicates, but systematic differences are clearly noticeable. Batch 1 shows good correlation across all replicates. Batches 2 and 3 show systematic deviations of both the diploid and high fitness lineages. Some of the differences in batch 2 explained by the low coverage time points in batch 2, replicate 2. The low coverage leads the inference to be dominated by a single slope. Batch 3 replicate 3 looks systematically different from the other replicates in batch even at high coverage time points.

### Cross-Batch Fitness Correlation

We next examine the correlations between batches, and found that these were worse than within-batch correlations. Figure S5 shows fitness-fitness correlations between the best replicates in each batch. (This is in contrast with Figure 2 in the main text, which compares the averages over each batch.)

### Systematic Differences between Batches for Specific Mutation Classes

Both the diploid and high fitness lineages exhibited systematic differences across batches. While the between replicate deviations were in the 1–2%/generation range, the between batch differences were as high as 5%/generation.

The last panel in Figure S5 compares the fitnesses across replicates and batches for the *GPB2*, *PDE2*, and diploid classes. Estimated error bars from the fitness assay are plotted. The fitnesses within a batch correlate well, with most deviations occurring in replicates with low coverage time points. The overall batch-batch systematics are different for different types of mutations. For most pairs of classes, the relative ordering does not change. However, some like *PDE2* and *GPB2* switch order in the different batches. These differences suggest that the systematic deviations are not merely an artifact of the fitness estimation algorithm and thus cannot be consistently corrected for statistically.

Within the best replicates, there is a very narrow spread of all but one of the *GPB2* mutant lineages, and all but one of the *PDE2* lineages. This suggests that the intrinsic precision and potential accuracy of the barcode fitness assay is  $\leq 1\%$ .

### Testing for Differences in Fitness Effect between Mutant Classes

The systematic cross-batch differences informed how we tested for differences in fitness effects between different mutation classes. We first carried out a number of ANOVA tests, for differences between genes, mutation types, and paralogs.

To test if gene identity was at all significant, we treated the batch as a categorical variable and still ended up with a  $P < 10^{-16}$ . For our tests of fitness difference of particular pairs, we carried out tests separately for each batch. We averaged over all time points and replicates within a batch to get a single fitness per lineage.

Within each batch we tested the hypothesis that the means of the fitness distributions for different classes were different. We found significant differences in fitness between *IRA1* and *GPB1*, *IRA1* and *IRA2*, *GPB1* and *GPB2*, and found that mutation type made a significant difference in the fitness of *IRA1* mutants (test was not significant for mutation type in *IRA2*). For the diploids, we found that a third copy of chromosome 11 gave significant fitness benefit, but a third copy of chromosome 12 did not. Additional coding mutations did not significantly change the fitness of diploids.

We also tested the null hypothesis that the distribution of *IRA1* mutants was the same as the distributions of *PDE2* and *GPB2* mutants. We used the non-parametric Kolmogorov-Smirnov (KS) test to test for any difference between distributions. The KS test compares the CDFs of two empirical distributions, and compares the largest gap between them (which is distributed in a way independent of distribution). The fact that the data pass the KS tests as well give us confidence that our results are not due to noise-modeling assumptions.

The results are also robust to changes in the fitness inference algorithm. If we instead use a weighted log-linear regression, choosing  $\hat{s}$  by

$$\hat{s} = \arg \min_s \sum_i \frac{(\log(r_{i+1}/R_{i+1}) - \log(r_i/R_i) + \mu_i - s)^2}{\kappa_i/r_{i+1}} \quad (18)$$

the fitness estimates change by  $\sim 1\%$  per generation at most. The differences in distributions and relative orderings of fitnesses persist.

## DATA AND SOFTWARE AVAILABILITY

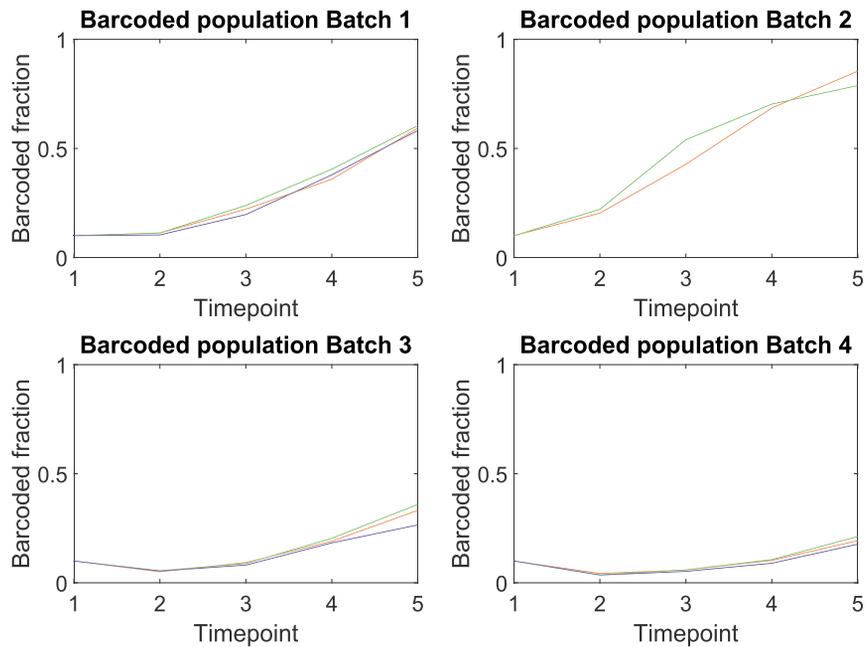
### Data Resources

All Illumina sequencing data (for both the whole-genome sequencing and the fitness measurement assays) can be found under NIH BioProject: PRJNA310010.

### Software

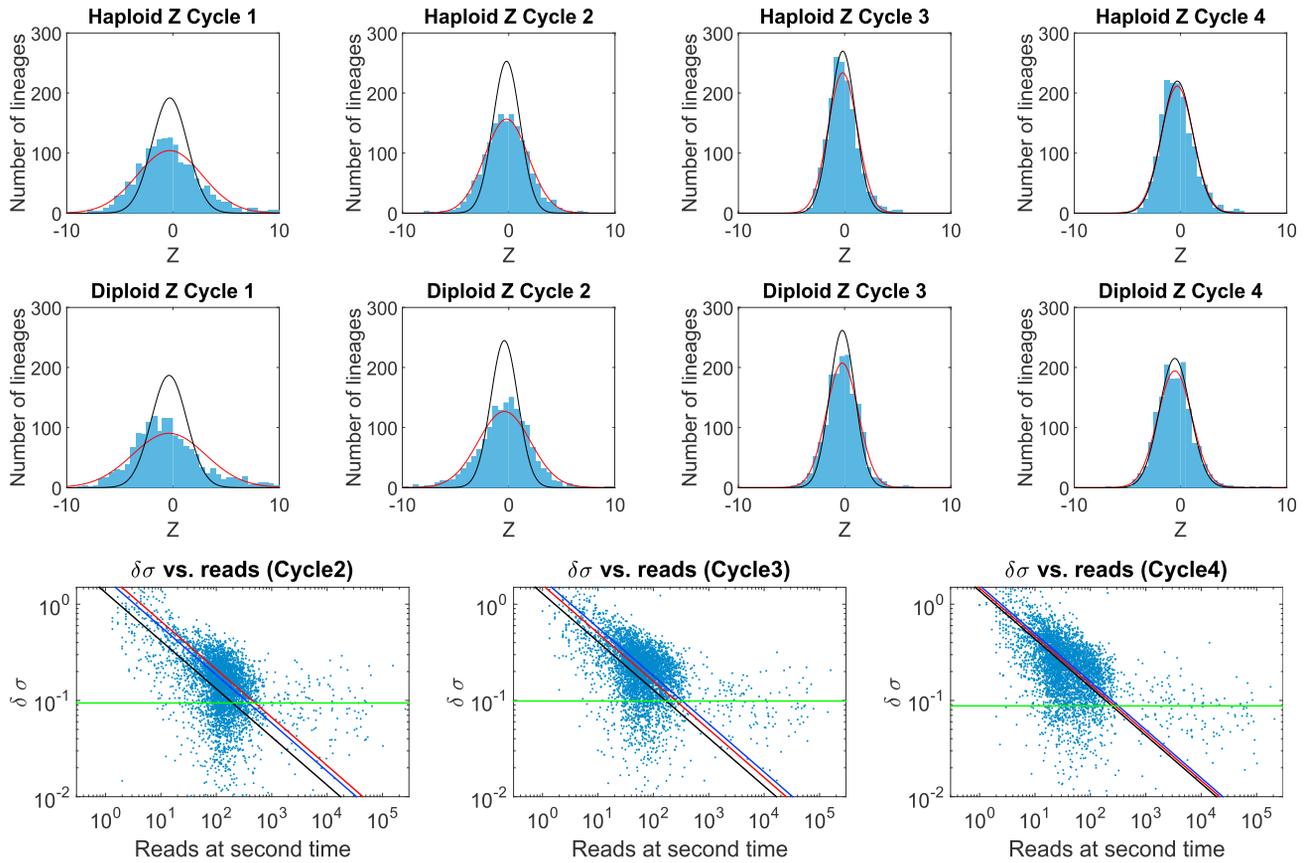
The software repository for the barcode counting code can be found at <https://github.com/sunthedeep/BarcodeCounter>.

# Supplemental Figures



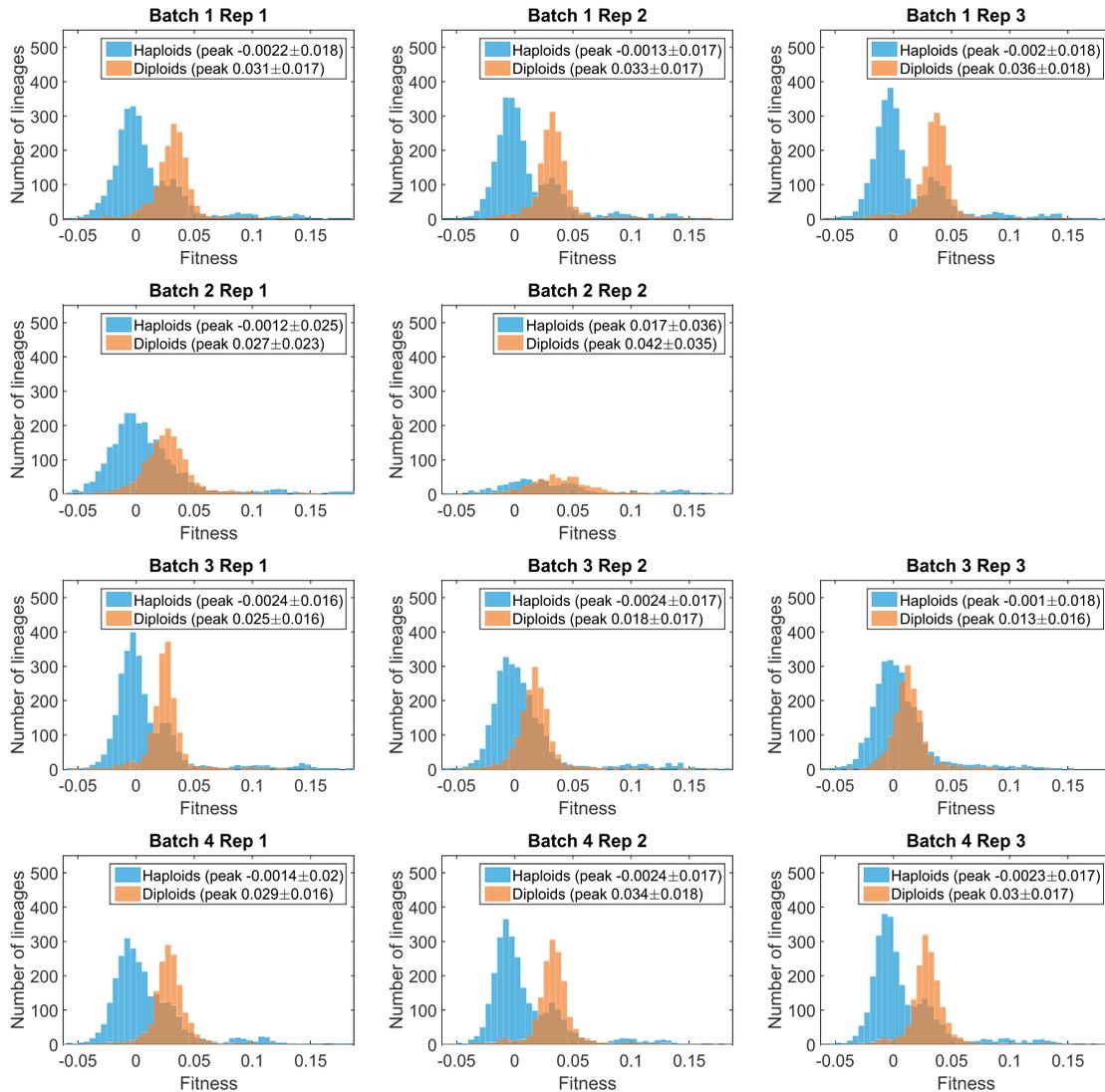
**Figure S1. Estimated Barcoded Fraction of Populations during Pooled Fitness Measurement, Related to Figure 2**

By the later time points, barcoded lineages have expanded considerably. Note that timepoint 1 was not used for fitness calculations for Batches 1, 3 and 4; however, the barcoded fraction is still around 10% at timepoint 2 for these batches.



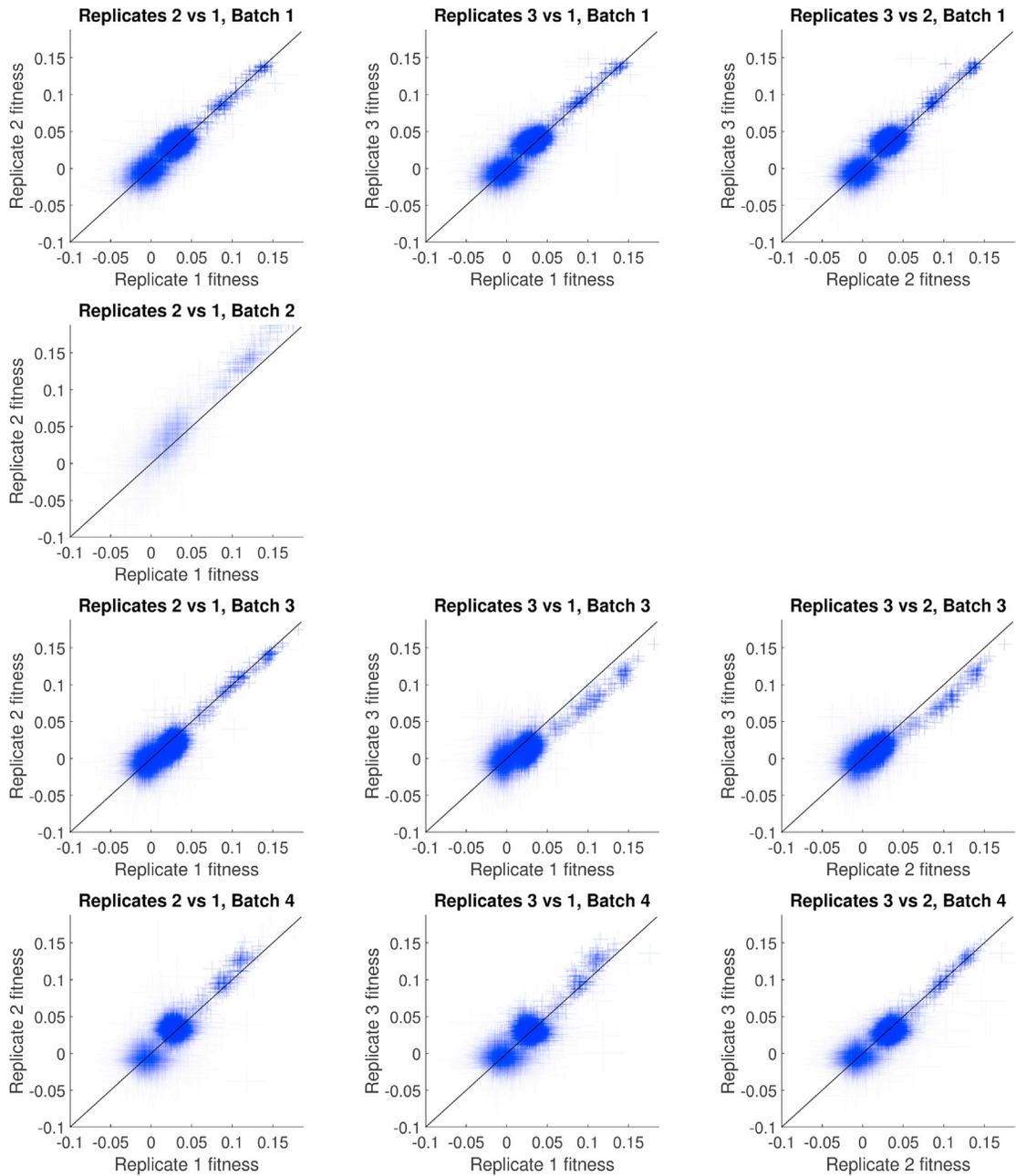
**Figure S2. Noise Analysis for Batch 1, Related to Figure 2**

The first two rows of panels show histograms of scaled deviations  $Z_i$  for haploid and diploid lineages over all cycles in replicate 1, including the predicted distribution from the noise model (red) and counting noise limit (black). Typical read counts are 100 at the beginning of the experiment and 30 by the end. The last row shows the SD of log-slope  $\delta\sigma_i$  vs. read number across all replicates. We expect the variance to scale as  $r^{-1}$ . Scaling estimates are shown from within replicate (red), between replicate (blue), and counting noise (black). We find that noise is independent of read count at high frequency (green) and approaches the minimal limit at late times.



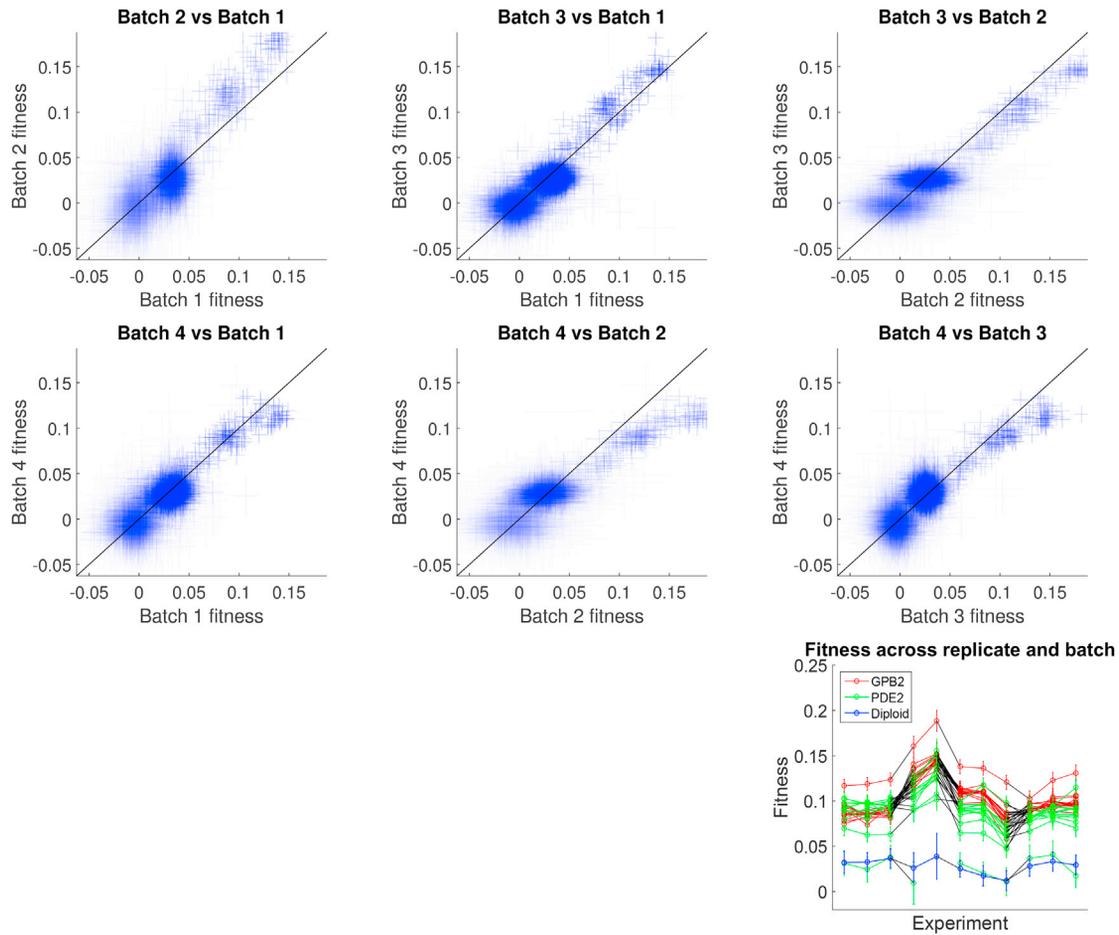
**Figure S3. Fitness Histograms for Barcoded Lineages, Related to Figure 2**

Each row corresponds to the replicates in a batch. We plot haploids in blue, and known diploids in red. We find that the haploid-diploid separation is poorly resolved in some experiments. Note that some fitness estimates in batch 2 replicate 2 failed due to low coverage at one of the timepoints.



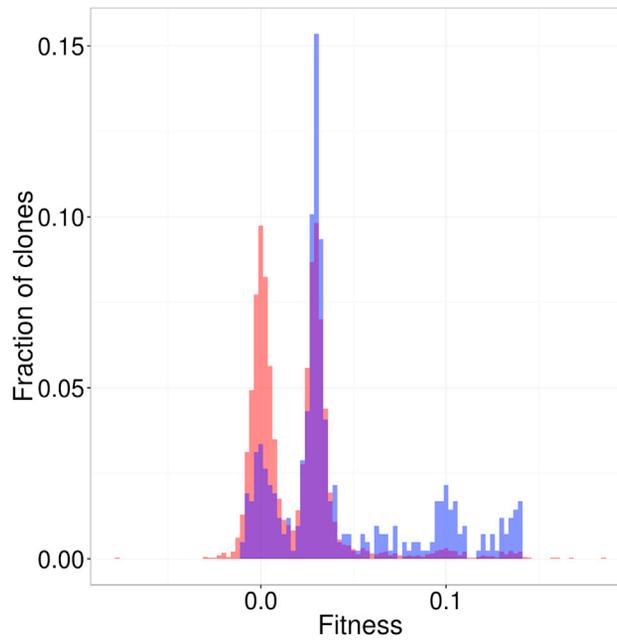
**Figure S4. Replicate-Replicate Fitness Comparisons, Related to Figure 2**

Batch 1 displays the highest consistency between replicates, while in Batch 3 both diploids and high-frequency lineages show systematic deviations.



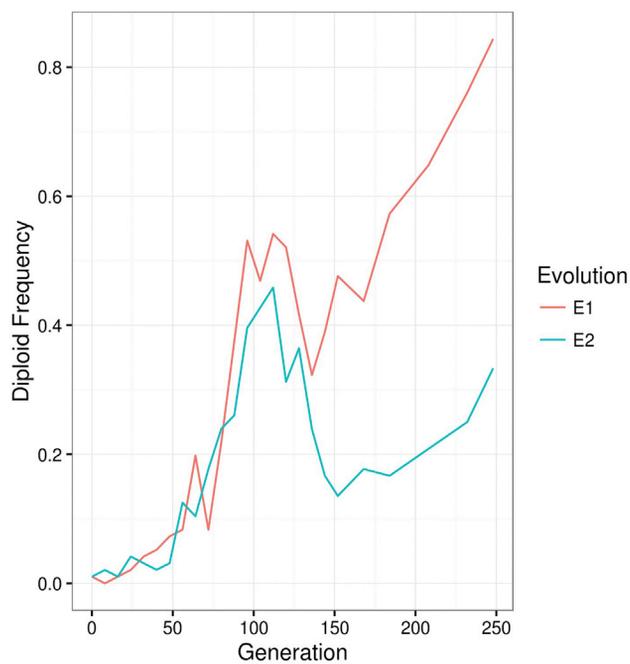
**Figure S5. Batch Effects in Fitness Measurements, Related to Figure 2**

The first 6 panels show batch-batch fitness comparisons. Here we show the best replicate from each batch to highlight deviations in the best of our measurements. Note that this is slightly different from Figures 2C and 2D, where we do the batch-batch comparisons using the weighted average fitness across replicates in the batch. Systematic divergences are at a scale of 1%–2% (per generation) for diploids, and larger for higher fitness lineages. The last panel shows fitnesses for GPB2, PDE2, and diploid classes with the batches separated by black bars. We found that within replicate variation is in the 1%–2% range, while the variation between batches can be as large as 5%. Systematic variation differs for different classes.



**Figure S6. Distribution of Fitness Effects of the 418 Sequenced Clones Compared to that of the 4,800 Randomly Sampled Clones, Related to Figures 1 and 4**

We intentionally over-enriched for high-fitness lineages among our 418 whole-genome sequenced clones (fitness distribution shown in blue) as compared to the fitness distribution of the 4,800 sampled clones (red), in order to sample as many adaptation-driving mutations as possible.



**Figure S7. Frequency of Diploids throughout the Levy et al. Replicate Evolutions, Related to Table 1**

Using frozen stocks of the Levy et al. (2015) time points, for each of the two independent evolution experiments (E1 and E2), we assayed the proportion of the population that consists of diploid cells across the entire course of the evolution.