High-throughput analysis of adaptation using barcoded strains of *Saccharomyces cerevisiae*

Vincent J. Fasanello*, Ping Liu^, Carlos A. Botero†[a], Justin C. Fay‡^[a]

* Division of Biology and Biomedical Sciences; Washington University School of Medicine in St. Louis; St. Louis, Missouri; 63110

^Department of Genetics; Washington University in St. Louis; St. Louis, Missouri; 63110

† Department of Biology; Washington University in St. Louis; St. Louis, Missouri; 63130

‡ University of Rochester; Rochester, New York; 14627

[a]These authors share senior authorship

Running Title: Barcoded Experimental Evolution

Keywords: Experimental Evolution, *Saccharomyces cerevisiae*, Genetic Barcode, Fitness Assay

Corresponding Author: Vincent J. Fasanello

ABSTRACT

Background: Experimental evolution of microbes can be used to empirically address a wide range of questions about evolution and is increasingly employed to study complex phenomena ranging from genetic evolution to evolutionary rescue. Regardless of experimental aims, fitness assays are a central component of this type of research, and low-throughput often limits the scope and complexity of experimental evolution studies. We created an experimental evolution system in *Saccharomyces cerevisiae* that utilizes genetic barcoding to overcome this challenge.

Results: We first confirm that barcode insertions do not alter fitness in this system and can be used to efficiently detect fitness differences via pooled competition-based fitness assays. Next, we examine the effects of ploidy, chemical stress, and population bottleneck size on the evolutionary dynamics and fitness gains (adaptation) in a total of 76 experimentally evolving populations by conducting 2,136 fitness assays and analyzing 532 longitudinal-evolutionary samples collected from the evolving populations. In our analysis of these data we describe the strengths of this experimental evolution system and explore sources or error in our measurements of fitness and evolutionary dynamics.

Conclusions: Our experimental treatments generated distinct fitness effects and evolutionary dynamics, respectively quantified via multiplexed fitness assays and barcode lineage tracking. These findings demonstrate the utility of this new resource for designing and improving high-throughput studies of experimental evolution. The approach described here provides a framework for future studies employing experimental designs that require high-throughput multiplexed fitness measurements.

BACKGROUND

Experimental evolution in microorganisms such as yeast, bacteria, and viruses has been used to answer evolutionary questions that are experimentally intractable in organisms with longer generation times [1–4]. A central benefit of most experimental evolution systems is the ability to replicate and repeat evolution as well as the ability to store and compete evolved, intermediate, and ancestral strains [4]. Indeed, if there is a single unifying theme to what we have learned from experimental evolution it is that adaptation is universal and often repeatable due to parallel changes down to the molecular level [5–9]. The power of this approach has led to numerous investigations of the effects of population size [6,10,11] and structure [12–14], mutation rate [15–18], and various environmental treatments [19–24].

Measuring adaptation by changes in fitness is a core requirement of experimental evolution that often limits its implementation. Fitness assays involve direct competition of individuals or populations against one another or a common reference, and traditionally have been calculated using neutral markers scored by plating assays. For example, the long-term experimental evolution conducted by Lenski and colleagues is scored by counting colonies that can ferment arabinose based on colony color [25]. Many recent studies use fluorescently marked strains such that the fitness of co-cultured strains can be assayed by flow-cytometry [e.g., 10,26,27]. However, the number of fitness assays, and thereby the resolution of those assays, remain limited, which poses a challenge for large-scale projects. The primary constraint limiting the scale of these studies is that measuring fitness requires replicate assays, often under a variety

67 of conditions. For example, an experiment with 100 strains evolved in a single environment
68 would require 800 fitness assays if one were to measure, in quadruplicate, the fitness of each
69 ancestral strain and each evolved strain in relation to a common reference. Furthermore, each
70 of the fitness assays would require the frequency of the two strains to be quantified at the
71 beginning and end of the competition in order to calculate fitness, resulting in 1600 individual
72 samples for a comprehensive look at fitness change in the 100 strains in question. Equally
73 important is the ability to detect small changes in fitness, which is directly related to the
74 number of replicate assays per strain and the noise of the assay itself.
75
76 Genetic barcodes can vastly increase the throughput of experimental evolution analyses
77 through pooled-fitness assay designs. In a barcoding approach, each strain is marked by
78 insertion of a unique neutral DNA sequence (i.e., a barcode) into its genome, enabling the easy
79 quantification of the relative abundance of multiple strains when simultaneously competed
80 against a common reference. Microarrays [28] and more recently, direct sequencing of
81 barcodes [29], have been used to measure the effects of thousands of single gene deletions on
82 fitness using this approach. Barcodes have also been used for lineage tracking in experimental
83 evolution. For example, Blundell and co-workers utilized a pool of half a million barcoded yeast
84 strains to detect and track the fate of lineages in large populations [30,31]. A further advantage
85 of this approach is that the barcodes can be used to measure the fitness of strains that have
86 been isolated from an evolved population in a traditional competition assay [32,33].
87

88 In this study we describe a set of barcoded yeast strains and characterize their utility for
89 assessing adaptation during experimental evolution. The system is composed of a collection of
90 *Saccharomyces cerevisiae* strains, individually barcoded with unique 20 bp sequences inserted
91 upstream of the *HO* locus; this barcoding strategy offers a number of advantages for
92 experimental evolution. First, cross-contamination between populations with different
93 barcodes can be detected and monitored. Second, populations can be initiated with mixtures of
94 multiple barcodes for tracking adaptive dynamics that occur during the experimental evolution
95 period [27,34]. Finally, fitness can be measured from the entire set of strains in a single pooled
96 fitness assay, which dramatically increases efficiency. To demonstrate the capabilities of this
97 system we began by conducting a series of proof-of-concept fitness assays and subsequently
98 applied what we learned in a short, 25-day (~250-generation) experimental evolution. These
99 analyses confirm that barcode insertions (1) do not alter the fitness of our source strains, and
100 (2) provide a high-throughput means of measuring fitness differences and evolutionary
101 dynamics for individual lineages from pooled samples obtained at different stages of
102 experimental evolution. We highlight the advantages of multiplexing samples with indexing as
103 well as the importance of limiting molecular contamination between initial sampling and library
104 construction. Taken together, this system represents a valuable resource for designing and
105 improving high-throughput studies in experimental evolution.
106
107 METHODS
108
109 **Strains, media and culture methods**

110 Barcoded yeast strains were constructed using two haploid derivatives of a diploid strain
111 collected from an Oak tree in Pennsylvania (YPS163) [35]: YJF153 (MATa, *HO*::dsdAMX4) and
112 YJF154 (MATalpha, *HO*::dsdAMX4). Barcoded kanMX deletion cassettes were amplified from
113 the MoBY plasmid collection [36] with primers containing homology to the promoter region (-
114 1,129 to -1,959) of *HO*. A set of 184 barcoded cassettes were selected based on confirmation of
115 correct barcodes by sequencing (Additional file 1 contains entries for barcodes used in this
116 project), then transformed into YJF153 and confirmed by PCR. Barcoded diploid strains were
117 made by mating barcoded haploids (YJF153) to YJF154, and confirming diploids by mating-type
118 PCR [37]. The final barcoded yeast strain library consisted of 184 haploid and 184 diploid
119 strains; 92 strains from this set were utilized in the proof-of-concept fitness assays and 77
120 strains from the set were utilized in the 250-generation experiment. Strains were stored at -
121 80°C as 15% glycerol stocks.
122
123 All evolution and fitness assays were conducted in complete medium (CM;2 g/l dextrose, 0.17
124 g/l yeast nitrogen base without amino acid and ammonium sulfate, 0.5 g/l ammonium sulfate,
125 0.13 g/l dropout mix complete without yeast nitrogen base) with or without additional stresses
126 in 96-deep well plates (2.2-ml poly-propylene plates, square well, v-conical bottom; Abgene AB-
127 0932) covered with rayon acrylate breathable membranes (Thermo Scientific, 1256705).
128 Growth plates were incubated at 30°C for 24 hours inside an incubator (VWR, Forced Air
129 Incubator, basic, 120v, 7 cu. ft.) with agitation using a horizontal electromagnetic microplate
130 shaker (Union Scientific LLC, 9779-TC). Saturated 24-hour culture was diluted (1:1000) into

131 fresh medium at the same time each day to initialize the next round of growth for all evolution
132 and fitness assays.
133
134 Starting material for all evolution and fitness assays originated from -80°C freezer stocks of the
135 constructed barcoded yeast strains. Yeast were revived from -80°C freezer stocks via a single
136 round of growth (1 day, 10 generations) under standard culture conditions for these assays.
137 Samples collected during our experiments were stored both as (1) 15% glycerol stocks at -80°C
138 to maintain viable freezer stocks of yeast populations [i.e., a "frozen time vault", sensu 4], and
139 (2) pelleted samples at -20°C for DNA extraction.
140
141 **Experimental design**
142 **Proof-of-concept fitness assays**: The design of this new system for experimental evolution
143 began with a proof-of-concept analysis that allowed us to optimize our methods and to confirm
144 that the fitness of multiple strains in pooled samples could be simultaneously and accurately
145 measured with sequencing-based competition assay methods. This initial step involved
146 measuring the fitness of 91 barcoded yeast strains relative to an ancestral reference strain
147 simultaneously using a sequencing-based fitness assay (Figure 1, A.). Ten replicate fitness assays
148 were conducted under standard culture conditions (see *Strains, media and culture methods*).
149 Briefly, 92 yeast strains were revived from -80°C freezer stocks, mixed in equal proportions (i.e.,
150 such that each strain comprised 1/92 of the pooled population), and diluted (1:1000) into fresh
151 medium to initialize the two-day proof-of-concept fitness assays. Samples were obtained from
152 the undiluted initial mixtures and, 20 generations later, from the final overnight population

153 cultures. Fitness was measured by the change in barcode abundance relative to a 'reference'
154 strain (d1H10) from 10 replicates over the two-day period of approximately 20 generations, for
155 a total of 920 fitness assays (92 focal barcodes x 10 replicates). See *Fitness calculations*, below,
156 for a full description of the competition-based fitness assay methodology and for calculations of
157 fitness from barcode abundance data. DNA was isolated separately for each fitness assay
158 sample using a ZR Fungal/Bacterial DNA Kit (Zymo Research D6005) in individual 2.0 mL screw-
159 cap tubes following the manufacturer's instructions. Physical cell disruption by bead-beating
160 was carried out in a mixer mill (Retsch, MM 300) at 30 Hz (1800 min^{-1}) for ten minutes (1-
161 minute on, 1-minute off, times ten cycles). Following extraction, DNA was amplified with
162 forward/reverse primers containing a 9-12 bp index for multiplex sequencing. PCR products
163 were quantified, pooled and purified to form a single multiplexed library for sequencing.
164 Additional control samples were also included in the library to track barcode cross-
165 contamination. See *Library construction and sequencing*, below, for a detailed description of
166 the library preparation protocol used for these and all other samples.
167
168 **250-Generation evolution experiment**: After establishing the feasibility of the analysis strategy
169 and optimizing culture, assay, and processing methods, 152 yeast strains were evolved for 25
170 days (i.e., ca. 250 generations at 9.97 generations per day) under different scenarios of
171 selection in a second set of experiments. Specifically, these yeast strains were evolved by serial
172 dilution in one of six different treatments (Figure 1, B1.; See Additional file 2 for treatment
173 descriptions). Evolutionary treatments involved growth in either complete media (CM), CM
174 with ethanol (8% by volume) or CM with NaCl (0.342 M). Serial transfers were achieved either

175 through standard dilution (1:1000), reduced dilution (1:250), or increased dilution (1:4000).
176 Note that the 1:250 dilution treatment and the 1:4000 dilution treatment are expected to have
177 undergone 200 and 300 generations of evolution, respectively, rather than the 250 generations
178 of evolution expected in the treatments with a standard 1:1000 dilution. This difference in
179 number of generations is caused by variation in the number of possible doubling events when
180 different numbers of cells are added to media with the same (limiting) amount of glucose.
181 Additionally, haploid and diploid yeast were evolved under standard culture conditions (see
182 *Strains, media and culture methods*) to assess the effects of ploidy. To initialize this evolution
183 experiment, barcoded yeast strains were revived from -80°C freezer stocks. Barcoded yeast
184 pairs slated to evolve in competition were then mixed in equal proportions and diluted in fresh
185 medium, according to the experimental design, and allowed to evolve for 25 days.
186
187 Samples were obtained from the initial undiluted mixtures and on day 25 (ca. 250 generations
188 later) to use, respectively, as the starting material for the Generation-0 and the Generation-250
189 fitness assays . These end-point assays of relative fitness were developed to characterize fitness
190 outcomes in the 250-generation experiment (See *End-point assay of relative fitness change*).
191 Additional samples were obtained at regular intervals to characterize evolutionary dynamics
192 (See *Evolutionary dynamics*). These two sets of assays are described in detail in the two
193 following sections.
194
195 *Evolutionary Dynamics:* Relative proportions of barcoded yeast strain pairs evolved in
196 competition within the same well were quantified at generations 0, 100, 150, 200, 220, 240,

197 and 250 (Days 0, 10, 15, 20, 22, 24 and 25) for a total of 532 evolutionary dynamics samples
198 across 76 evolving two-barcode populations. These samples were pooled for DNA extraction
199 such that there was no barcode overlap within pools. DNA was subsequently extracted (see the
200 *Proof-of-concept fitness assays* section for details). Libraries were constructed for sequencing as
201 described in the *Library construction and sequencing* section. From these data, the relative
202 proportions of barcoded yeast strains within each pair were measured at each time-point.
203 Using this information, the time-point (t-max) and magnitude (m-max) of the maximum change
204 in relative abundance in comparison to the starting conditions, the time-point (t-max-rate) and
205 magnitude (m-max-rate) of the maximum rate of change between adjacent time-points, the
206 time-point (t-max-diff) and maximum difference (m-max-diff) in barcode proportions, and the
207 total cumulative change in barcode relative abundance were calculated for each two-barcode
208 population. Barcodes approaching fixation (hereafter referred to as "fixed") were also noted
209 and were defined as cases in which a single barcode in the pair obtained (and maintained) a
210 proportion of 0.95 or greater by (through) generation 250 of the evolution experiment.
211
212 *End-point assays of relative fitness:* This second type of assay was designed to assess the fitness
213 of a given focal strain relative to a reference using a barcoded competition-based assay.
214 Because they involve taking samples (and quantifying fitness) at the start and end of an
215 experimental evolution, end-point fitness assays provides a means of measuring fitness change
216 across the entire experiment in the design utilized here. In this case, the assays involved
217 quantifying barcode fitness using pooled samples from generation 0 (Figure 1, B2.) and
218 generation 250 (Figure 1, B3.) from the 250-generation evolution experiment (Figure 1, B1.).

219 Briefly, Generation-0 and Generation-250 yeast strains were revived and samples from each
220 time point were pooled such that there was no overlap in barcoded yeast strain identity within
221 each pool. Pools of revived generation-0 and generation-250 yeast were independently
222 combined in equal proportion (50 pooled-yeast : 50 ancestral reference) with an ancestral
223 reference strain (re: an 'unevolved' barcoded yeast strain); these pooled samples were then
224 diluted into fresh medium to initialize the fitness assays. Four replicate Fitness assays were
225 conducted for each pool of generation-0 and generation-250 yeast. Fitness assays employed a
226 standard 1:1000 transfer dilution across all samples (see *Strains, media and culture methods*).
227 Samples evolved in CM plus additional stresses were assayed in the same media type that they
228 were evolved in. Diploid yeast were competed against a diploid reference strain (strain ID:
229 d1H10), while haploid yeast were competed against the haploid version of this same reference
230 (strain ID: h1H10). In these assays, yeast samples for DNA extraction were obtained from the
231 undiluted initial mixtures (fitness assay starting material), and, 20 generations later, from the
232 final overnight population cultures (fitness assay end) to assess fitness of generation-0 and
233 generation-250 yeast strains. See *Fitness Calculations*, below, for a full description of fitness
234 assay methodology and for calculations of fitness from fitness assay barcode abundance data.
235 Libraries were constructed for sequencing as described in the *Library construction and*
236 *sequencing* section, below.
237
238 **Library construction and sequencing**: Barcode sequencing libraries for the proof-of-concept
239 fitness assays and all samples from both components of the 250-generation evolution
240 experiment were constructed by amplification of MoBY barcodes [36] with primers containing

241 Ion Torrent adaptors with indexes assigned to distinguish samples from one another (Additional
242 file 3).
243
244 PCR products for library construction were generated using 25 cycles and were subsequently
245 quantified with a Qubit 3.0 Flourometer, high sensitivity assay kit. These products were then
246 combined at equimolar concentrations and purified using a Zymo DNA Clean & Concentrator kit
247 (Zymo Research D4014) to create a single library for sequencing. The DNA extraction and PCR
248 steps were repeated for samples that did not attain sufficient DNA concentrations (no samples
249 were exposed to > 25 rounds of PCR). We note that PCR jackpotting can be an issue in DNA
250 library construction and that it can add noise to measurements of barcode frequency. To
251 minimize the influence of this phenomenon on our results, we used a large amount of template
252 DNA [38], and conducted all fitness assays in four replicates. Additionally, we avoided
253 misclassification of barcodes by using high fidelity Taq Plus DNA Polymerase (Lambda Biotech,
254 Taq Plus Master mix Red), and MOBY barcode sequences that differ from one another in at
255 least six substitutions.
256
257 DNA libraries were sequenced using an Ion Torrent sequencer (Ion Proton System, Ion Torrent)
258 at the Genomics Core Facility at Saint Louis University with a customized parameter to assess
259 polyclonality after 31bp (the start position of the forward Ion Torrent adapter index sequence).
260 A single sequencing run was used for each pooled library (library 1 – proof-of-concept fitness
261 assays, library 2 – evolution experiment: evolutionary dynamics samples, Generation-0 fitness

262 assays, and Generation-250 fitness assays). An additional library was constructed and
263 sequenced for a set of samples from library 2 with elevated barcode cross-contamination rates.
264
265 **Sequence data processing & calculations**
266 **Sequence datasets:** Sequence data in FASTQ format were parsed and demultiplexed using
267 custom scripts in R (see the *Availability of data and materials* statement, below). A total of
268 142,243,245 raw reads that matched the forward Ion Torrent adapter indices included in our
269 experiment (omitting reads that matched no forward adapter, polyclonal reads, low quality
270 reads, and adapter dimer reads) were recovered across the three sequenced libraries.
271 104,365,740 reads (73.4%) were retained for analysis that perfectly matched a forward
272 sequencing adapter index (9-12 bp), reverse sequencing adapter index (9-12 bp) pair, and a
273 MOBY genetic barcode (20 bp) included in the full experimental design.
274
275 **Barcode cross-contamination rate:** The barcoded yeast experimental evolution system has an
276 innate ability to detect and track barcode cross-contamination that could arise during
277 evolution, over the course of short-term fitness assays, or during DNA library preparation.
278 Barcode cross-contamination rate was defined as the mean number of counts mapping to any
279 given barcoded yeast strain included in the full experimental design (sequenced library), but not
280 expected to be present in the given sample (pair of forward and reverse IonTorrent adapter
281 Indices). Barcode cross-contamination rates were calculated separately for each unique
282 forward-reverse index pair included in each sequencing library and represent the amount of

283 noise an average contaminating barcode strain contributes to each sample. The barcode cross-

284 contamination rate of sample (primer pair) j, B_j, was measured as,

285

286 Equation 1.

$$B_j = \frac{\frac{1}{m}\sum_{i=1}^{m} C_i}{T_j}$$

287

288 where *m* is the number of barcoded strains that could potentially contribute to barcode cross-

289 contamination in experiment *j* (i.e., the number of unique strain IDs included in the full library,

290 but not expected in sample j), C_i is the number of barcoded yeast strain counts recovered for

291 contaminating barcode *i*, and T_j is the total number of counts recovered in sample *j* across all

292 barcoded yeast strains included in the full experimental design.

293

294 Contamination rate summaries are reported separately for the proof-of-concept fitness assays

295 and the evolution experiment (the latter containing evolutionary dynamics, generation-0 fitness

296 assay and generation-250 fitness assays samples). For the subset of evolution experiment

297 samples that were sequenced in two separate libraries, only the replicate with a lower

298 contamination rate was retained for contamination rate summary reporting and all

299 downstream analyses. In all statistical analyses reported below, contamination rate is initially

300 included as a potential predictor and subsequently removed from the model if its effect was

301 found to be non-significant.

302

303 **Fitness calculations**: Fitness was evaluated via sequencing-based assays that involved

304 competing barcoded yeast strains against a common ancestral reference strain for 2 days (20

305 generations). Reported fitness values for the yeast strains in all fitness assays are relative to the

306 same barcoded ancestral reference strain (strain ID: d1H10 for diploids; strain ID: h1H10 for

307 haploids). The relative fitness of barcoded yeast strain *i* at generation *gn,* $w_{i\,gn}$, was measured

308 as,

309

310 Equation 2.

$$w_{i_{gn}} = \left(\left(ln\frac{Ci_{20}}{R_{20}} - ln\frac{Ci_0}{R_0}\right)/20\right)^e$$

311

312 where C*i* and *R* refer to barcode counts for the focal barcode and reference barcode at fitness

313 assay generation 0 (initial mixtures; fitness assay initial measurement) and fitness assay

314 generation 20 (final overnight cultures; fitness assay final measurement), and *20* is the number

315 of generations between measurement at fitness assay generation 0 and fitness assay

316 generation 20 [39]. The change in fitness of strain *i* between Day 0 and Day 25 in for our 250-

317 generation evolution experiment, Δw_i, we therefore computed as,

318

319 Equation 3.

320 $$\Delta w_i = w_{i_{g250}} - w_{i_{g0}}$$

321

322 where $w_{i\,g0}$ and $w_{i\,g250}$ are the strain's fitness at generation 0 and 250 as measured from

323 Equation 2.

324

325 **Statistical analysis**

326 **Analysis & visualization tools:** All Analyses and statistical test were conducted in R version 3.5.2
327 [40]. Data processing uses primarily base-R functionality; the plyr package [41] is used in some
328 cases for data frame manipulation. All Statistical models with linear mixed effects were
329 generated using the lme4 [42], and lmerTest [43] packages. Weighted t-tests were assessed
330 with the weights package [44]. Power analyses were conducted using the pwr package [45].
331 Finally, figures and tables were generated with the ggplot2 [46] and sjPlot [47] packages,
332 respectively; multi-panel figures were built using the gridExtra package [48]. Raw p-values are
333 reported unless otherwise noted; some tables included in the supplement report raw and
334 corrected p-values (i.e., false discovery rate) for cases where only corrected values are
335 presented in the main text.

336

337 **Reads:** Analysis of multiplex barcode sequencing data requires careful consideration of sample
338 size. Because count data are ultimately handled as relative frequencies (proportions), it was
339 necessary to consider underlying sample size or "confidence" in each piece of data within the
340 full dataset for all calculations and analyses. That is, entries with more reads were explicitly
341 assumed to contribute more to summary calculations and statistical models. Variation in
342 sample size was thus controlled for by weighting all calculations by the read sample size and by
343 including such weights in downstream statistical models. This sample size metric considers both
344 the total counts recovered for a multiplexed sample (unique forward and reverse index
345 sequence adapter pair) and the number of counts recovered for the focal barcoded yeast

346 strain(s) for that entry. In the analyses presented here, read calculations for constructed
347 variables utilize harmonic means rather than arithmetic means; the harmonic mean is sensitive
348 to small values and therefore does not overinflate the read values for any variables constructed
349 from a combination of high- and low-read data sources.
350
351 **Proof-of-concept fitness assays:** Fitness differences among 92 constructed barcoded yeast
352 strains used in the proof-of-concept fitness assays were assessed using a weighted linear model
353 with mean-corrected fitness as the response variable and yeast strain ID as the predictor
354 variable. Barcode cross-contamination is not included in this model as all fitness assay samples
355 contained all 92 focal barcodes; barcode cross-contamination is reported from a separate set of
356 1-barcode and 2-barcode samples for the Proof-of-concept fitness assays.
357
358 **Power Analysis:** Power analyses were conducted to calculate the sensitivity of the system in
359 measuring treatment-level fitness effects, fitness changes for individual barcoded yeast strains,
360 and initial fitness variation among individual strains. Power was calculated using sample size
361 and fitness effect as variables of interest with a fixed value for error in fitness measurements.
362 Error in fitness was estimated from the root mean squared error among replicates from the
363 Proof-of-Concept fitness assays as well as from replicate measures of fitness change for
364 individual barcodes from the 250-generation experiment. Power analyses were conducted
365 using the pwr package in R; the pwr.f2.test function was used to calculate power to discern
366 treatment-level differences in fitness and the pwr.t.test function was used to calculate power

367 to detect individual barcode fitness changes and initial fitness differences among individual
368 barcodes.
369
370 **Multiplex barcode sequencing and cross-contamination:** Barcode cross-contamination rate
371 was assessed (Equation 2) and summarized, separately for the proof-of-concept fitness assays
372 and the 250-generation experiment. The effects of barcode cross-contamination rate on the
373 change in fitness and the error among replicate fitness measures for each barcode were
374 assessed by including barcode cross-contamination as a predictor in two sets of linear mixed-
375 effects models with either change in fitness over 250 generations of experimental evolution or
376 standard error in fitness change over 250 generations (across four replicate measures for each
377 barcode) as the response variable. Additionally, a weighted 1-tailed t-test was utilized to assess
378 whether barcode cross-contamination rates for samples collected in the 250-generation
379 experiment decreased after re-extracting DNA and resequencing.
380
381 **Fitness change in 250 generations of experimental evolution:** Individual barcoded yeast strains
382 that exhibited any change in fitness over the 250-generation fitness experiment were identified
383 using a weighted linear model with change in fitness as the response variable and a strain
384 identifier (treatment plus yeast strain Barcode ID) as the predictor variable.
385
386 Next, the effects of evolutionary treatment (medium type, ploidy, and transfer dilution) on
387 change in fitness over 250 generations of experimental evolution was assessed using a
388 weighted linear mixed-effects model with cross-contamination rates as a covariate. A random

389 effect of strain identifier (treatment plus yeast strain barcode ID) was placed on the model
390 intercept. A second weighted linear model was fit to assess potential sources of error in the
391 measurement of fitness in this system. Here, a model was constructed using a comprehensive
392 set of factors that could have impacted the standard error in fitness measurement among
393 replicates. These included: treatment, cross-contamination rate, magnitude fitness change,
394 median count (median number of counts returned for a focal barcode across all sampling points
395 that contribute to the fitness change calculation), and median proportion (median proportion
396 of the focal barcode in its mixed 2-barcode well across all sampling points).
397
398 **Evolutionary dynamics:** For all analyses of evolutionary dynamics, only the first barcoded yeast
399 strain from each well was included to ensure that data were not double-counted. A set of
400 weighted linear mixed effects models, with treatment as the predictor variable and either t-
401 max, m-max, t-max-rate, m-max-rate, t-max-diff, m-max-diff, or total cumulative change in
402 barcoded relative abundance across all time-points as the dependent variable, were employed
403 to assess treatment differences in evolutionary dynamics. Full models also included initial
404 barcode abundance as a predictor term because initial barcode abundance could impact
405 subsequent dynamics. Initial barcode abundance was subsequently removed from models
406 when deemed nonsignificant, resulting in removal from all but one model (t-max-diff). Barcode
407 fixation rate was not assessed statistically due to the small number of fixation events observed
408 (n=7/76 populations).
409
410 RESULTS

411

412 **Proof-of-concept fitness assays**

413 We constructed 92 diploid yeast strains, each with a unique 20 bp barcode inserted upstream
414 of the deleted *HO* gene. Proof-of-Concept fitness assays were then used to measure any fitness
415 differences among these constructed "barcoded" strains, to estimate our power to detect small
416 fitness differences, and to assess our ability to measure fitness using multiplexed barcode
417 sequencing. We assayed fitness simultaneously for our pool of 92 barcodes by measuring
418 changes in barcode abundance relative to a 'reference' strain (barcode ID: d1H10) in 10
419 replicates over a two-day period of approximately 20 generations. A total of 920 fitness assays
420 were performed in this section of this study (92 barcodes x 10 replicates).

421

422 A few of the barcoded strains showed significant differences in fitness (3/92 at a 5% false
423 discovery rate, FDR; 1/92 at a 1% FDR) (Figure 2, Additional file 4). The root mean squared error
424 (rMSE) among replicated measures of fitness was 1.76e-2, indicating good power (80%) to
425 detect fitness differences of 1.2% between strains with four replicates at a nominal significance
426 level of P = 0.05 (Figure 3a, Additional file 5). We subsequently used these data to estimate
427 power to detect treatment effects in our proposed experimental design for the 250-generation
428 experiment. With a modest number of barcoded yeast strains per treatment (n=22), we
429 estimate high power (99.8%) to detect 1% fitness differences between treatments at a nominal
430 significance level of P = 0.05 with four replicate fitness assays (Figure 3b, Additional file 6). With
431 these promising results, we proceeded to conduct a more comprehensive test of the utility of

432 barcoded strains in a practical experimental evolution context (i.e., the utility to evaluate the
433 effects of treatment on fitness outcomes and evolutionary dynamics).
434
435 **Experimental evolution**
436 To empirically evaluate the strengths of this barcoded-strain system for studies of experimental
437 evolution, we conducted a 25-day evolution experiment, which is approximately equal to 250
438 generations of evolution. Our design included 76 populations, each initiated with two barcodes
439 per well. Samples spanned six treatments, which varied in yeast strain ploidy, growth medium,
440 and daily transfer dilution (Additional file 2). To measure any changes in fitness we competed
441 the evolved generation-250 strains and their generation-0 ancestors against a common
442 reference strain over a two-day period of approximately 20 generations. With four replicate
443 fitness assays for each barcode, this amounted to 1,216 fitness assays (2 barcodes * 2 time-
444 points * 76 populations * 4 replicates). We also measured the change in barcode frequency of
445 barcoded strains within each evolving population (barcode pair) from 532 longitudinal-
446 evolutionary dynamics samples (76 barcoded pairs * 7 timepoints). From these samples we
447 quantified the magnitude and timing of changes in barcode frequency, which should be
448 influenced by changes in the fitness of the evolving barcoded strains present within each
449 population.
450
451 **Multiplex barcode sequencing**: We utilized a two-step multiplexed design to obtain high
452 throughput estimates of fitness based on barcode sequencing in our fitness assays. In the first
453 step we leveraged the strain-identifying barcodes by pooling multiple strains together and

454 simultaneously competing them against an ancestral reference strain to estimate relative
455 fitness. In our second multiplexing step we PCR-amplified each fitness assay sample with unique
456 forward and reverse indexed sequencing adaptors. This latter step enabled us to assign
457 sequencing reads to the appropriate fitness assay after sequencing many samples in concert as
458 a single library. We constructed one library for each experiment (Library 1 – Proof-of-Concept
459 fitness assays; Library 2 – 250-generation evolution experiment evolutionary dynamics and
460 fitness assays samples).
461
462 From 84,776,627 demultiplexed reads, we obtained a median of 2,961 reads per barcode in
463 each sample. However, we also detected barcode cross-contamination, defined as barcodes
464 with sample indexes that shouldn't exist in our sequenced libraries. The average rate of
465 barcode cross-contamination per sample was low, 0.04%, and consistently present in nearly all
466 of our samples (Additional file 7 A, B). Contamination could occur during culturing of yeast
467 strains, liquid handling during preparation of libraries [4,25], or potentially via index switching
468 during sequencing procedures [49–51]. However, the low but uniform rate of cross
469 contamination we observe is more consistent with library preparation and sequencing than
470 yeast contamination. This is further evidenced by the fact that no growth was observed in
471 culture blanks during any fitness assays nor during the 25-day experimental evolution
472 experiment.
473
474 To test whether the barcode cross-contamination rate depends on liquid handling during library
475 preparation, we reprocessed a subset of samples with high cross contamination rates using

476 identical starting material (multiple yeast sample aliquots were created at the time of sample
477 collection). Cross contamination rates decreased significantly in these reprocessed samples ($t = 22.3$, $df = 65$, $p < 10e-6$), but a low level of background contamination remained (Original
478
479 Samples: mean = 0.75%; reprocessed samples: mean = 0.05%) (Additional file 8).
480
481 In our analyses, the presence of low abundance barcode cross-contamination was removed
482 from all samples in which a barcode is not expected to occur. However, this doesn't eliminate
483 contamination in samples where a barcode is expected to occur. In such cases, error in
484 estimates of barcode frequency is highest when a barcode's frequency is low and approaches
485 the cross-contamination rate. The effects of contamination rate and other system-specific
486 covariates on experimental outcomes are presented in detail in the next section.
487
488 **End-point fitness assays**: We assessed fitness in the 152 evolved strains (evolution experiment
489 generation-250 strains) as well as their ancestors (evolution experiment generation-0 strains).
490 For each strain and timepoint, fitness was measured in comparison to a common ancestral
491 reference strain via a two-day competition-based fitness assay in the same media type under
492 which the strain had been evolved. The root mean squared error (rMSE) among replicated
493 measures of fitness was 4.31e-2 for these assays. This is slightly higher than our proof-of-
494 concept fitness assays but is still expected to provide reasonable power to detect fitness
495 changes for individual barcodes (80% power to detect fitness changes of 2.99%) and fitness
496 differences among treatments (80% power to detect treatment differences of 0.81%) (Figure 3c
497 & Figure 3d, Additional file 9, Additional file 10). Among a number of covariates tested we

498 found that only the cross-contamination rate (P = 0.01) and magnitude fitness change (P =
499 6.03e-4) were associated with error among replicates (Additional File 11).
500
501 Twenty three percent (35/152) of the strains showed significant increases in fitness between
502 generation 0 and generation 250 at a 1% false discovery rate, FDR (Figure 4, Additional File 12).
503 Within this subset, the average fitness increase was 6.80% with a range of 2.75% to 23.5%.
504 Relatively few strains with significant increases in fitness were found in the CM treatment
505 (3/42; 7.14%) and ethanol stress treatment (2/22; 9.09%). A higher proportion of strains
506 exhibited significant increases in fitness in the lower 1:250 dilution treatment (9/22; 40.9%), the
507 salt stress treatment (11/22; 50.0%), and the haploid treatment (10/22; 45.5%). No strains in
508 the 1:4000 dilution treatment (0/22) exhibited increases in fitness. No strains (0/152) exhibited
509 a significant decrease in fitness.
510
511 Next, we evaluated the effect of evolutionary treatment on fitness change. We found
512 significant variation in fitness among treatments (P < 10e-6, Figure 5, Additional file 13).
513 Relative to our standard culture conditions, diploids in CM ("no stress") with 1:1000 transfer
514 dilution, we found greater (more positive) fitness change in diploid strains evolved in salt stress
515 (P < 10e-6), in haploid strains evolved in standard culture conditions (P = 0.01), and in diploid
516 strains evolved with a less extreme (1:250) daily transfer dilution (P = 9.49e-4). Relative to the
517 CM ("no stress") diploid treatment (our standard culture conditions), we found less (less
518 positive) fitness change in diploid strains evolved in ethanol stress (p = 3.99e-3). We found no
519 significant effect of a more extreme (1:4000) daily transfer dilution on fitness change (p=0.39).

520 In addition to the treatment effects, there was also a negative effect of barcode cross
521 contamination rate on fitness change (P < 10e-6).
522
523 **Evolutionary dynamics**: In experimental evolution, adaptation can influence the relative
524 abundance of barcodes evolving in competition [27,34]. In addition to characterizing fitness
525 change for the evolved barcodes, we tracked the relative proportions of the 76 barcoded-yeast
526 pairs at seven timepoints during our 250-generation evolution experiment (days 0, 10, 15, 20,
527 22, 24, and 25) to examine the evolutionary dynamics generated by adaptation and other
528 processes (e.g., drift). We collected 8 measurements from these 532 evolutionary dynamics
529 samples (76 populations * 7 timepoints): (1) barcode fixation, (2) the time-point and (3)
530 magnitude of the maximum change in relative abundance in comparison to the starting
531 conditions, (4) the time-point and (5) magnitude of the maximum rate of change between
532 adjacent time-points, (6) the time-point and (7) magnitude of the maximum difference in BC
533 proportions, and (8) the total cumulative change in barcoded relative abundance summed
534 across all time-points.
535
536 Out of the seven measures of adaptive dynamics that were amenable to statistical testing (i.e.,
537 all adaptive dynamics measures other than fixation rate, which had sparse imbalanced binary
538 data), six significantly differed by treatment (Table 1; see additional files 14:27 for individual
539 adaptive dynamics plots and results tables). Relative to the CM-diploid treatment (our standard
540 culture conditions) both the salt and haploid treatments were associated with a larger
541 maximum change in relative abundance in comparison to the start (p = 5.24e-5, p = 5.85e-3,

542 respectively) and a more extreme maximum difference in barcode proportion (p = 2.77e-4, p =
543 6.01e-6, respectively). The salt treatment was also associated with an earlier time-point of the
544 maximum change in abundance relative to the start (p = 0.01), and the haploid treatment was
545 associated with an earlier maximum rate of change (p = 1.35e-5). The total cumulative change
546 in barcode abundance was greater for the 1:4000 dilution treatment than the 1:1000 dilution
547 treatment (p = 1.77e-3). Total cumulative change in barcode abundance was significantly less
548 for haploids than diploids (p = 4.9e-2), and significantly less in the ethanol stress treatment than
549 in the CM with no stress treatment (p = 2.98e-2). Finally, we observed barcodes that
550 approached fixation in eleven percent (9/76) of our populations. Most fixation events were
551 observed in the salt (3/11; 27.3%) and haploid (5/11; 45.5%) treatments, one instance was
552 found in the CM-diploid treatment and no instances of fixation were observed in the ethanol,
553 1:4000 dilution, nor 1:250 dilution treatments.
554
555 DISCUSSION
556
557 Experimental evolution has proven to be a valuable approach for studying a range of
558 evolutionary questions. In this study we implemented a genetic barcoding system in *S.*
559 *cerevisiae* to increase the efficiency and throughput of these types of studies. Through
560 barcoding of ancestral strains we were able to increase the complexity of our experimental
561 design, increase the throughput of fitness assays by multiplexing of independent samples, and
562 provide a relatively simple means to track the evolutionary dynamics of barcodes evolving in
563 competition during the experimental evolution period. Accordingly, we have demonstrated the

potential for detecting fitness differences in six experimental evolution treatments and have shown that, while the typically low levels of barcode cross-contamination we observe cannot be completely eliminated, their effects on inference can be minimized through simple statistical procedures. Below we discuss the merits and drawbacks of the barcoding system, and its capabilities to increase efficiency and throughput in measuring fitness changes that occur during experimental evolution.

One important consideration of a barcoding system like the one we examine here, is that barcoded strains are not necessarily identical to one another at the beginning of an experiment (i.e., have equal fitness) even if all barcoded variants are produced from a single ancestral clone. Although we found no significant fitness differences among the majority of barcoded strains, we note that we did indeed observe a few strains with significant deviations from the population mean fitness. Given the location of our barcode insertions (i.e., a non-functional region of the genome) it is unlikely that the barcodes themselves generated these fitness differences. A perhaps more likely explanation is that these differences arose from mutations that occurred during transformation [29] or shortly thereafter. Regardless, this limitation could be mitigated by (1) removing strains with unequal generation-0 fitness entirely from their analyses, (2) quantifying initial fitness differentials and including them as a covariate in downstream analyses, or (3) looking explicitly at change in fitness for each individual barcode as we do here. We note that initial fitness differentials may be of interest themselves, given that they can potentially impact evolutionary outcomes [52–54] and the types of mutations that successfully spread through an experimental population [55].

586

587 In addition to throughput, the power to detect small changes in fitness in evolved strains is a
588 critical parameter of any experimental evolution system. We estimated 80% power to detect a
589 fitness difference of 1.2% in the proof-of-concept assays and 80% power to detect a fitness
590 change of 3.0% in the 250-generation evolution experiment for any individual barcode. While a
591 formal power analysis is not often reported, a review of the literature suggests that prior
592 studies have detected fitness differences between 1 and 5% [26,32,56,57]. For example, one
593 previous barcoded fitness assay found between replicate deviations of 1-2% and between batch
594 deviations of up to 5% [32]. Some studies report confidence intervals significantly lower than
595 1%, but these studies tend to employ a large number of replicates [12,55,58]. We also note that
596 our power to detect treatment effects was much higher than our ability to detect changes in
597 fitness for individual barcoded strains. Specifically, we estimated 99.8% power to detect fitness
598 differences of 1% between treatments using the error in our proof-of-concept fitness assays,
599 and 87.7% power to detect fitness differences of 1% between treatments using the actual 250-
600 generation experiment data.

601

602 While the error measured in our fitness assays is consistent with most prior studies, pooled
603 fitness assays do have additional sources of error. Prior studies using pooled fitness assays
604 estimated that pools containing 1000 cells per strain are expected to have an error rate of 4.4%
605 [59] after two rounds of dilution due to sampling error. The expected number of cells per
606 barcoded strain at inoculation in our system is 9,000 (3.0x10^8 cells/ml at saturation in 0.6ml of
607 media, diluted 1:1000 and divided by up to 20 strains per assay). However, the number of cells

608 could be much lower due to sampling or to low frequency of the barcode of interest in a well at
609 the end of the 250 generations of evolution. Lineages with large changes in fitness are expected
610 to deviate the most from equal numbers of cells in a well. Consistent with this possibility, the
611 RMSE of fitness in generation 250 (4.34) is higher than the RMSE of fitness in generation 0
612 (2.92), where barcode frequencies are much more uniform. This provides one explanation for
613 our finding that error among replicates is associated with the magnitude of the fitness change.
614 Thus, while there are some limitations to pooled fitness assays, these limitations can be
615 compensated by examination a large number of strains per treatment –as done in this study–
616 or by increasing the number of replicate assays.
617
618 Another important consideration when designing an experimental evolution system is that it
619 must be fairly robust to contamination. While culture contamination is rare in experimental
620 evolution [25,60], barcode cross-contamination is possible. Our results are consistent with prior
621 work on this topic. Specifically, while we found no evidence of culture contamination in any
622 growth blanks, we detected a uniformly low level of diffuse barcode cross-contamination in
623 multiplexed fitness assays and evolutionary dynamics samples. Because there was no
624 conspicuous pattern of cross-contamination, we suggest that the observed contamination was
625 likely introduced during sample preparation and/or DNA sequencing. DNA extraction is a likely
626 source of cross-contamination in samples processed in strip-tubes or 96-well plate formats that
627 prioritize throughput. However, we minimized the chance for contamination in the DNA
628 isolation step in our experiments by isolating DNA with individual reaction tubes for each
629 sample. Furthermore, contamination in the DNA extraction step would not be consistent with

630 the low-level diffuse contamination observed, a pattern that tends to be more indicative of
631 contamination occurring downstream of DNA extraction. Several other sources of
632 contamination are possible, including primer contamination during the index addition for PCR
633 [61] and index switching during library construction or sequencing [49–51]. The latter possibility
634 seems nevertheless less likely because all PCR steps were performed separately prior to pooling
635 of libraries and because no previous studies report index switching for IonTorrent sequence
636 data to our knowledge.
637
638 The strength and efficiency of barcoded experimental evolution is further evidenced by this
639 study. Barcoding enabled us to evolve a large number of strains (152) for 250 generations
640 across six treatments, and to conduct a total of 532 evolutionary dynamics measurements and
641 1,216 fitness assays related to these manipulations in a relatively short amount of time and
642 with limited resources. Reassuringly, our results are largely consistent with prior work. We find
643 greater fitness increases, i.e., a greater rate of adaptation, in haploids than diploids. This finding
644 agrees with another study that found faster rates of adaptation and larger effective population
645 sizes in haploids relative to diploids [10]. In a related study, Selmecki et al., [27] found faster
646 adaptation in tetraploids than diploids, but no difference in rate of adaptation between
647 haploids and diploids, potentially suggesting a trend opposite to ours of increasing adaptive
648 rate with increasing ploidy. However, differences between haploids and diploids must be
649 treated with caution because there is mounting evidence that haploid yeast evolves towards
650 diploidy under experimental evolution conditions [20,27,62]. We also find greater fitness
651 increase in complete medium (CM) plus NaCl stress than in CM alone, which was not surprising

652 given what is known about adaptation to NaCl stress in *S. cerevisiae* [20,63–65]. In contrast, we
653 were surprised to detect less fitness increase in CM plus EtOH stress than in CM alone. There
654 are several non-mutually exclusive explanations for this result. It is possible that ethanol did not
655 present a significant stress (selective pressure) to the cells once they had attained physiological
656 adaptation to the medium, i.e., acclimation [19]. It is also possible that adaptations to CM and
657 adaptations to ethanol exhibit antagonistic pleiotropy, similar to what has been found in
658 experiments contrasting rich and poor media [66] or exploring adaptation to other chemical
659 stressors [67]. Pleiotropy could also shed light on the marked adaptation observed in the NaCl
660 treatment given that adaptation to CM and NaCl stress may exhibit complementarity via
661 synergistic or positive pleiotropy [21,68–70]. We find no difference in fitness change between
662 our standard (1:1000) dilution treatment and a treatment with a more extreme (1:4000) daily
663 transfer dilution. We do, however, find a greater increase in fitness when a less extreme (1:250)
664 daily transfer dilution was used instead of the standard (1:1000) dilution. While these transfer
665 dilution findings are not necessarily expected [12,71–73], they nevertheless support earlier
666 findings that less extreme bottlenecks favor the maintenance of adaptive mutants [74] and that
667 large populations are less adaptively constrained than small ones in simple environments [73].
668 They are also consistent with previous evidence for greater mean fitness increase in wide
669 versus narrow bottleneck populations [11].
670
671 In addition to high-throughput fitness assays, barcoding enabled us to track the relative
672 proportions of barcoded lineage pairs evolving in competition during the experimental
673 evolution period [30,31,75,76]. As expected, we found general agreement between the

674 evolutionary dynamics results and endpoint fitness assay results in our six experimental
675 treatments. For example, the haploid and NaCl stress treatments both displayed dynamics
676 consistent with more extreme increases in fitness, including greater change in barcode
677 abundance relative to the starting conditions and a greater maximum difference in barcode
678 abundance than diploids in CM. Interestingly, haploids also showed signs of earlier adaptation
679 than diploid strains in similar conditions, as evidenced by an earlier generation of maximum
680 rate of change in barcode abundance (also noted in [27,30,31]) and a lower total change in
681 barcode abundance over 250 generations. Although this latter result may seem paradoxical, it is
682 consistent with the observation that haploids adapt earlier than diploid strains [10], and is
683 expected if a greater proportion of the total change in barcode abundance of haploids in our
684 experiments happened in the first 100 generations (before we collected any evolutionary
685 dynamics data).
686
687 Despite significant differences in barcode dynamics, there are some limitations to interpreting
688 these results. Because we mostly assessed abundance every 50 generations, it is possible that
689 we missed some of the adaptive dynamics; dense temporal sampling is ideal for a full picture of
690 evolutionary dynamics in populations subjected to experimental evolution [76]. Furthermore,
691 due to projected read depth constraints, barcode frequencies over time were not measured in
692 replicate for the 250-generation evolution project. Finally, changes in barcode abundance over
693 long time scales may not be directly related to final fitness as these measures are calculated
694 over 250 generations of evolution, during which relative fitness relationships between evolving
695 barcodes could change significantly. For example, the 1:4000 transfer dilution bottleneck

696 treatment had elevated rates of change in barcode abundance and a high amount of total
697 change in barcode abundance without a concomitant increase in fitness, possibly as a result of
698 the potentially strong effects of drift when severe bottlenecks are frequent. We suggest that
699 future studies employ a denser and more even longitudinal-evolutionary dynamics sampling
700 scheme, with replication, to maximize the value and precision of this type of lineage tracking or
701 evolutionary dynamics data.
702

703 CONCLUSIONS

704

705 In summary, we conclude that the barcoded yeast system that we describe here greatly
706 increases the throughput of fitness measurements and provides a relatively simple means for
707 lineage tracking. These two characteristics greatly increase our ability to design more complex
708 and potentially more useful evolutionary experiments. We note that although barcode cross-
709 contamination imposes some limitations on the implementation of this system, it is possible to
710 track the origin and rates of such contamination and, therefore, to statistically consider its
711 effects on experimental outcomes. Thus, we conclude that this system represents a significant
712 step forward toward the design and implementation of experimental evolution systems capable
713 of high throughput fitness measurements.
714

715 REFERENCES

716

717 1. de Varigny H. Experimental Evolution. London, United Kingdom: MacMillan and Co; 1892.

718 2. Garland T, Rose MR. Experimental Evolution concepts, methods, and applications of
719 selection experiments. University of California Press; 2009.
720 3. Kassen R. Experimental Evolution and the Nature of Biodiversity. Greenwood Village, CO
721 USA: Roberts and Company publishers, Inc.; 2014.
722 4. Van den Bergh B, Swings T, Fauvart M, Michiels J. Experimental Design, Population
723 Dynamics, and Diversity in Microbial Experimental Evolution. Microbiol Mol Biol Rev
724 [Internet]. 2018;82(3):e00008-18. Available from:
725 http://mmbr.asm.org/lookup/doi/10.1128/MMBR.00008-18
726 5. Bailey SF, Guo Q, Bataillon T. Identifying drivers of parallel evolution: A regression model
727 approach. Genome Biol Evol. 2018;10(10):2801–12.
728 6. Bailey SF, Blanquart F, Bataillon T, Kassen R. What drives parallel evolution?: How
729 population size and mutational variation contribute to repeated evolution. BioEssays.
730 2017;39(1):1–9.
731 7. Graves JL, Hertweck KL, Phillips MA, Han M V., Cabral LG, Barter TT, et al. Genomics of
732 parallel experimental evolution in drosophila. Mol Biol Evol. 2017;34(4):831–42.
733 8. Burke MK, Liti G, Long AD. Standing genetic variation drives repeatable experimental
734 evolution in outcrossing populations of saccharomyces cerevisiae. Mol Biol Evol.
735 2014;31(12):3228–39.
736 9. Kohn LM, Anderson JB. The underlying structure of adaptation under strong selection in
737 12 experimental yeast populations. Eukaryot Cell. 2014;13(9):1200–6.
738 10. Gerstein AC, Cleathero LA, Mandegar MA, Otto SP. Haploids adapt faster than diploids
739 across a range of environments. J Evol Biol. 2011;24(3):531–40.

740 11. Schoustra SE, Bataillon T, Gifford DR, Kassen R. The properties of adaptive walks in
741 evolving populations of fungus. PLoS Biol. 2009;7(11).
742 12. Kryazhimskiy S, Rice DP, Desai MM. Population subdivision and adaptation in asexual
743 populations of saccharomyces cerevisiae. Evolution (N Y). 2012;66(6):1931–41.
744 13. Low-Décarie E, Kolber M, Homme P, Lofano A, Dumbrell A, Gonzalez A, et al. Community
745 rescue in experimental metacommunities. Proc Natl Acad Sci. 2015;112(46):14307–12.
746 14. Bell G, Gonzalez A. Adaptation and Evolutionary Rescue Environmental Deterioration.
747 Science (80-) [Internet]. 2011;332(June):1327–30. Available from:
748 http://www.ncbi.nlm.nih.gov/pubmed/21659606
749 15. Loewe L, Textor V, Scherer S. High Deleterious Genomic Mutation Rate in Stationary
750 Phase of Escherichia coli. Science (80-). 2003;302(5650):1558–60.
751 16. Swings T, van Den Bergh B, Wuyts S, Oeyen E, Voordeckers K, Verstrepen KJ, et al.
752 Adaptive tuning of mutation rates allows fast response to lethal stress in escherichia coli.
753 Elife. 2017;6:1–24.
754 17. Lenski RE, Sniegowski P, Gerrish P. Evolution of high mutation rates in experimental
755 populations of E. coli. Nature. 1997;387(June):703–5.
756 18. Perfeito L, Fernandes L, Mota C, Gordo I. Adaptive mutations in bacteria: High rate and
757 small effects. Science (80-). 2007;317(5839):813–5.
758 19. Huang CJ, Lu MY, Chang YW, Li WH. Experimental Evolution of Yeast for High-
759 Temperature Tolerance. Mol Biol Evol. 2018;35(8):1823–39.
760 20. Dhar R, Sägesser R, Weikert C, Yuan J, Wagner A. Adaptation of Saccharomyces
761 cerevisiae to saline stress through laboratory evolution. J Evol Biol. 2011;24(5):1135–53.

762 21. Dhar R, Sägesser R, Weikert C, Wagner A. Yeast adapts to a changing stressful
763 environment by evolving cross-protection and anticipatory gene regulation. Mol Biol
764 Evol. 2013;30(3):573–88.
765 22. Hughes BS, Cullum AJ, Bennett AF. Evolutionary adaptation to environmental pH in
766 experimental lineages of Escherichia coli. Evolution (N Y). 2007;61(7):1725–34.
767 23. Horinouchi T, Suzuki S, Hirasawa T, Ono N, Yomo T, Shimizu H, et al. Phenotypic
768 convergence in bacterial adaptive evolution to ethanol stress. BMC Evol Biol [Internet].
769 2015;15(1):1–14. Available from: http://dx.doi.org/10.1186/s12862-015-0454-6
770 24. Zhou A, Baidoo E, He Z, Mukhopadhyay A, Baumohl JK, Benke P, et al. Characterization of
771 NaCl tolerance in Desulfovibrio vulgaris Hildenborough through experimental evolution.
772 ISME J [Internet]. 2013;7(9):1790–802. Available from:
773 http://dx.doi.org/10.1038/ismej.2013.60
774 25. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-Term Experimental Evolution in
775 Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. Am Nat
776 [Internet]. 1991;138(6):1315–41. Available from:
777 http://www.journals.uchicago.edu/doi/10.1086/285289
778 26. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, et al. The repertoire and
779 dynamics of evolutionary adaptations to controlled nutrient-limited environments in
780 yeast. PLoS Genet. 2008;4(12).
781 27. Selmecki AM, Maruvka YE, Richmond PA, Guillet M, Shoresh N, Sorenson AL, et al.
782 Polyploidy can drive rapid adaptation in yeast. Nature [Internet]. 2015;519(7543):349–
783 51. Available from: http://dx.doi.org/10.1038/nature14187

784 28. Roth FP, Smith AM, Kaper F, Nislow C, Giaever G, Thompson MJ, et al. Quantitative
785 phenotyping via deep barcode sequencing. Genome Res. 2009;19(10):1836–42.
786 29. Giaever G, Nislow C. The yeast deletion collection: A decade of functional genomics.
787 Genetics. 2014;197(2):451–65.
788 30. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. Quantitative
789 evolutionary dynamics using high-resolution lineage tracking. Nature.
790 2015;519(7542):181–6.
791 31. Blundell JR, Levy SF. Beyond genome sequencing: Lineage tracking with barcodes to
792 study the dynamics of evolution, infection, and cancer. Genomics [Internet].
793 2014;104(6):417–30. Available from: http://dx.doi.org/10.1016/j.ygeno.2014.09.005
794 32. Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, et al. Development of a
795 Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. Cell.
796 2016;166(6):1585-1596.e22.
797 33. Li Y, Venkataram S, Agarwala A, Dunn B, Petrov DA, Sherlock G, et al. Hidden Complexity
798 of Yeast Adaptation under Simple Evolutionary Conditions. Curr Biol [Internet].
799 2018;28(4):515-525.e6. Available from: https://doi.org/10.1016/j.cub.2018.01.009
800 34. Kao KC, Sherlock G. Molecular characterization of clonal interference during adaptive
801 evolution in asexual populations of Saccharomyces cerevisiae. Nat Genet.
802 2008;40(12):1499–504.
803 35. Sniegowski PD, Dombrowski PG, Fingerman E. Saccharomyces cerevisiae and
804 Saccharomyces paradoxus coexist in a natural woodland site in North America and
805 display different levels of reproductive isolation from European conspecifics. FEMS Yeast

806 Res. 2002;1(4):299–306.

807 36. Magtanong L, Andersen RJ, Porter J, Barker SL, Nislow C, Boone C, et al. A molecular
808 barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. Nat
809 Biotechnol. 2009;27(4):369–77.

810 37. Huxley C, Green ED, Dunbam I. Rapid assessment of S. cerevisiae mating type by PCR.
811 Trends Genet. 1990;6(C):236.

812 38. Cha RS, Thilly WG. PCR Methods and Applications: Specificity, Efficiency, and Fidelity of
813 PCR. Genome Res [Internet]. 1993;S18–30. Available from:
814 https://genome.cshlp.org/content/3/2/S8.full.pdf

815 39. Hartl DL, Clark AG. Principles of Population genetics. Vol. 116. Sunderland, MA: Sinauer
816 associates; 1997.

817 40. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna,
818 Austria; 2012. Available from: http://www.r-project.org/

819 41. Hadley Wickham. The Split-Apply-Combine Strategy for Data Analysis. J Stat Softw
820 [Internet]. 2011;40(1):1–29. Available from:
821 http://www.ncbi.nlm.nih.gov/pubmed/12345678

822 42. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models using lme4. J
823 Stat Softw [Internet]. 2015;67(1). Available from: http://arxiv.org/abs/1406.5823

824 43. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: Tests in Linear Mixed
825 Effects Models . J Stat Softw. 2017;82(13).

826 44. Pasek J. Weights: Weighting and Weighted Statistics. 2018.

827 45. Champely S. pwr: Basic Functions for Power Analysis. 2018.

828 46. Wikham H. ggplot2: Elegant Graphics for Data Analysis. 1st ed. New York: Springer-Verlag
829 New York; 2009.
830 47. Ludecke D. sjPlot: Data Visualization for Statistics in Social Science [Internet]. 2019.
831 Available from: https://cran.r-project.org/package=sjPlot
832 48. Auguie B. gridExtra: Miscellaneous Finctions for "Grid" Graphics. 2017.
833 49. Sinha R, Stanley G, Gulati GSG, Ezran C, Travaglini KJK, Wei E, et al. Index Switching
834 Causes "Spreading-Of-Signal" Among Multiplexed Samples In Illumina HiSeq 4000 DNA
835 Sequencing. bioRxiv [Internet]. 2017;125724. Available from:
836 https://www.biorxiv.org/content/early/2017/04/09/125724
837 50. Costello M, Fleharty M, Abreu J, Farjoun Y, Ferriera S, Holmes L, et al. Characterization
838 and remediation of sample index swaps by non-redundant dual indexing on massively
839 parallel sequencing platforms. BMC Genomics. 2018;19(1):1–10.
840 51. Illumina. Effects of Index Misassignment on Multiplexing and Downstream Analysis.
841 Illumina. 2017;1–4.
842 52. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global Epistasis Makes Adaptation
843 Predictable Despite sequence-level stochasticity. Science (80-). 2014;344(6191):1519–
844 22.
845 53. Barrick JE, Kauth MR, Strelioff CC, Lenski RE. Escherichia coli rpoB mutants have
846 increased evolvability in proportion to their fitness defects. Mol Biol Evol.
847 2010;27(6):1338–47.
848 54. Jerison ER, Kryazhimskiy S, Mitchell JK, Bloom JS, Kruglyak L, Desai MM. Genetic variation
849 in adaptability and pleiotropy in budding yeast. Elife. 2017;6.

850 55. MacLean RC, Perron GG, Gardner A. Diminishing returns from beneficial mutations and
851 pervasive epistasis shape the fitness landscape for rifampicin resistance in Pseudomonas
852 aeruginosa. Genetics. 2010;186(4):1345–54.
853 56. Hughes BS, Cullum AJ, Bennett AF. An Experimental Evolutionary Study on Adaptation to
854 Temporally Fluctuating pH in *Escherichia coli*. Physiol Biochem Zool [Internet].
855 2007;80(4):406–21. Available from:
856 http://www.journals.uchicago.edu/doi/10.1086/518353
857 57. Barrick JE, Lenski RE. NIH Public Access. Spring. 2010;(517):119–29.
858 58. Gallet R, Cooper TF, Elena SF, Lenormand T. Measuring selection coefficients below 10 -3:
859 Method, Questions, and Prospects. Genetics. 2012;190(1):175–86.
860 59. Pierce SE, Davis RW, Nislow C, Giaever G. Genome-wide analysis of barcoded
861 saccharomyces cerevisiae gene-deletion mutants in pooled cultures. Nat Protoc.
862 2007;2(11):2958–74.
863 60. Cooper TF, Lenski RE. Experimental evolution with E. coli in diverse resource
864 environments. I. Fluctuating environments promote divergence of replicate populations.
865 BMC Evol Biol. 2010;10(11).
866 61. Lo YMD, Mehal WZ, Fleming KA. False-Positive Results and the Polymerase Chain
867 Reaction. Lancet. 1988;332(8612):679.
868 62. Gerstein AC, Otto SP. Cryptic fitness advantage: Diploids invade haploid populations
869 despite lacking any apparent advantage as measured by standard fitness assays. PLoS
870 One. 2011;6(12).
871 63. Park WK, Yang JW, Kim HS. Identification of novel genes responsible for salt tolerance by

872 transposon mutagenesis in Saccharomyces cerevisiae. J Ind Microbiol Biotechnol.
873 2015;42(4):567–75.
874 64. Tekarslan-Sahin SH, Alkim C, Sezgin T. Physiological and transcriptomic analysis of a salt-
875 resistant saccharomyces cerevisiae mutant obtained by evolutionary engineering. Bosn J
876 Basic Med Sci. 2018;18(1):55–65.
877 65. Blomberg A. Global changes in protein synthesis during adaptation of the yeast
878 Saccharomyces cerevisiae to 0.7 M NaCl. J Bacteriol. 1995;177(12):3563–72.
879 66. Minty JJ, Lesnefsky AA, Lin F, Chen Y, Zaroff TA, Veloso AB, et al. Evolution combined with
880 genomic study elucidates genetic bases of isobutanol tolerance in Escherichia coli.
881 Microb Cell Fact. 2011;10:1–38.
882 67. Reyes LH, Abdelaal AS, Kao KC. Genetic Determinants for n -Butanol Tolerance in Evolved
883 Escherichia coli Mutants: Cross Adaptation and Antagonistic Pleiotropy between n -
884 Butanol and Other Stressors. Appl Environ Microbiol. 2013;79(17):5313–20.
885 68. Ostman B, Hintze A, Adami C. Impact of epistasis and pleiotropy on evolutionary
886 adaptation. Proc R Soc B Biol Sci. 2012;279(1727):247–56.
887 69. McGee LW, Sackman AM, Morrison AJ, Pierce J, Anisman J, Rokyta DR. Synergistic
888 pleiotropy overrides the costs of complexity in viral adaptation. Genetics.
889 2016;202(1):285–95.
890 70. Hughes KA, Leips J. Pleiotropy, constraint, and modularity in the evolution of life
891 histories: insights from genomic analyses. Ann N Y Acad Sci. 2017;1389(1):76–91.
892 71. Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual
893 population. Genetica. 1998;102(107):127–44.

894 72. De Visser JAGM, Zeyl CW, Gerrish PJ, Blanchard JL, Lenski RE. Diminishing returns from
895 mutation supply rate in asexual populations. Science (80-). 1999;283(5400):404–6.
896 73. De Visser JAGM, Rozen DE. Clonal interference and the periodic selection of new
897 beneficial mutations in Escherichia coli. Genetics. 2006;172(4):2093–100.
898 74. Wahl LM, Gerrish PJ, Saika-Voivod I. Evaluating the impact of population bottlenecks in
899 experimental evolution. Genetics. 2002;162(2):961–71.
900 75. Cooper VS. Experimental Evolution as a High-Throughput Screen for Genetic Adaptations.
901 mSphere. 2018;3(3):1–7.
902 76. Hegreness M, Shoresh N, Hartl D, Kishony R. An Equivalence Principle for the
903 Incorporation of Favorable Mutations in Asexual Populations. 2006;311(March):1615–8.
904 77. Fasanello VJ, Liu P, Botero CA, Fay JC. Data From: High-throughput analysis of adaptation
905 using barcoded strains of Saccharomyces cerevisiae [Internet]. 2019. Available from:
906 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA555990/
907 78. Fasanello VJ. VinceFasanello/MM_Code_Supplement: Code Supplement v1.0.0: High
908 throughput analysis of adaptation using barcoded strains of Saccharomyces cerevisiae.
909 2020 Jan 24 [cited 2020 Jan 24]; Available from:
910 https://doi.org/10.5281/zenodo.3627318
911
912 DECLARATIONS
913
914 **Ethics approval and consent to participate**
915 Not applicable

916

917 **Consent for publication**

918 Not applicable

919

920 **Availability of data and materials**

921 The dataset supporting the conclusions of this article is available in the NCBI Sequence Read
922 Archive (SRA) repository, BioProject Number PRJNA555990,
923 https://www.ncbi.nlm.nih.gob.bioproject/PRJNA555990 [77]. Data formatted for analysis,
924 intermediate data frames, as well as the custom R scripts utilized for all data processing,
925 statistical analysis, and figure generation are available from GitHub
926 (github.com/VinceFasanello/MM_Code_Supplement); a static version of the repository is
927 available from Zenodo [78]. A readme file is available in the repository with the instructions
928 necessary to reproduce the analyses and to confirm the results presented in this article.
929 Supplementary figures, tables, and files referenced throughout the main text are available as
930 "additional files"; detailed descriptions of all such files can be found in Additional File S1. Strains
931 are available from the Fay Lab upon request.

932

933 **Competing interests**

934 The authors have no competing interests.

935

936 **Funding**

V.F. was supported by funding provided by the National Science Foundation Graduate Research Fellowship Program (NSF-GRFP) (DGE-1745038) and the Botero and Fay labs. J.C.F. was supported by the National Institutes of Health (GM080669).

Authors' contributions

J.C.F., C.A.B., P.L., and V.J.F conceived of and designed the study. P.L., and V.J.F. conducted the lab work. V.J.F. processed the data, performed the statistical analysis., and created visualizations. V.J.F., J.C.F., and C.A.B. drafted and edited the manuscript.

Acknowledgements

The authors thank the members of the Botero and Fay labs, specifically Angela Chira, Trevor Fristoe, Emery Longan, and Dilys Vela for ongoing feedback and helpful advice.

FIGURES

Figure 1

Overview of experimental design. Barcoded yeast strains were used for proof-of-concept fitness assays (A, 2 days) and experimental evolution (B, 25 days) with gray boxes indicating one day (10 generations). During the 25-day, 250-generation, experimental evolution (B1), evolutionary dynamics samples were collected at generations 0, 100, 150, 200, 220, 240, and 250 (cyan arrows). Fitness of the ancestral (B2) and evolved (B3) strains were quantified via (2-day, 20-generation) fitness assays (red arrows).

959

960 **Figure 2**

961 Histogram of fitness from 92 barcoded yeast strains. Fitness is the deviation from the
962 population mean and was quantified by competition against a common reference strain via
963 Proof-of-concept fitness assay. Orange and green bars indicate yeast strains with fitness values
964 significantly and not significantly different from the population mean at a false discovery rate of
965 5%, respectively.

966

967 **Figure 3**

968 Power surfaces for POC fitness assays data (green) and fitness change data from the 250-
969 generation experiment (orange). Power to detect individual barcode fitness deviation from
970 population mean calculated using error in the POC fitness assay data (A, green) and power to
971 detect individual barcode fitness change using error in the 250-generation experiment (A,
972 orange); both employ four replicate measures of fitness. Power to detect treatment-level
973 fitness differences for two treatments with 22 barcodes in each as estimated using error in the
974 POC fitness assay data (B, green) and the 250-generation experiment (B, orange).

975

976 **Figure 4**

977 Change in fitness over 250 generations of experimental evolution. Histogram depicts fitness
978 increase quantified by competition against a common reference strain. Fitness increase is
979 generation-250 fitness minus generation-0 fitness. Orange and green bars indicate strains with
980 fitness values significantly greater or not significantly greater than their own fitness at

generation 0 at a false discovery rate of 1%, respectively. No strains significantly decreased in fitness. 16/152 entries with large standard error (> 0.029) among replicates are not displayed, but are included for all statistical testing. Dashed lines indicate the means of each group, with black indicating the overall mean.

Figure 5

Change in fitness across six treatments in 250 generations of experimental evolution. Violin plots show the density of 152 yeast strains' change in fitness with treatment labels indicating the medium, ploidy and dilution rate. Points indicate mean fitness change across 4 replicates for individual barcodes; point sizes reflect the number of reads underlying each datapoint and colors indicate evolutionary treatments. Treatment mean fitness changes are depicted as heavy black crossbars. 16/152 entries with large standard error (> 0.029) among replicates are not displayed, but are included for all statistical testing. Treatments significantly different from the reference, diploid yeast evolved under a standard 1:1000 transfer dilution in CM, are marked with an asterisk.

1003 TABLES

1004

1005 **Table 1**

1006 Associations between evolutionary dynamics and evolutionary treatments.

	Treatment[1]				
Measure	**CM + ETOH**	**CM + NaCl**	**1:4000 dil.**	**1:250 dil.**	**Haploid**
t-max	N.S.	-48.702 **	N.S.	N.S.	N.S.
m-max	N.S.	0.292 ***	N.S.	N.S.	0.178 **
t-max-rate	N.S.	N.S.	N.S.	N.S.	-52.717 ***
m-max-rate	N.S.	N.S.	0.006 *	N.S.	N.S.
t-max-diff	N.S.	N.S.	N.S.	N.S.	N.S.
m-max-diff	N.S.	0.175 ***	N.S.	N.S.	0.204 ***
total change	-0.007 *	N.S.	0.018 *	N.S.	-0.007 *

1007

1008 [1]Estimates for significant differences between each evolutionary treatment and the control

1009 (CM-diploid, 1:1000 transfer) are shown for seven different measures of barcoded dynamics (t-

1010 max, m-max, t-max-rate, m-max-rate, t-max-diff, m-max-diff, and total change).

1011

1012 ADDITIONAL FILES

1013

1014 **File Name: Additional file 1**

1015 File Format: Word Document, .docx

1016 Title of Data: Yeast strains and inserted MOBY barcode sequences.

1017 Description of Data: Columns 1 and 2, resp., contain diploid and haploid yeast strain ID's. Only
1018 yeast strains used in the current project are included (diploids, n = 92; haploids, n = 13).
1019 Columns 3 and 4, resp., contain the MOBY barcode uptag (5'-3') and MOBY barcode downtag
1020 (3'-5-) sequences that uniquely identify the haploid and diploid yeast strains in the same row.
1021
1022 **File Name: Additional file 2**
1023 File Format: Word Document, .docx
1024 Title of Data: Treatments for 250-generation experimental evolutions.
1025 Description of Data: Treatment ID is a numerical ID for each of six treatments; Evo. Wells is the
1026 number of populations assigned to each treatment; Barcodes per Evo. Well is the number of
1027 sympatric barcodes per evolutionary population; Ploidy, Evo. Medium, and Evo. Transfer
1028 Dilution indicate the yeast strain ploidy, evolutionary medium type, and daily passage dilution
1029 for each treatment, respectively.
1030
1031 **File Name: Additional file 3**
1032 File Format: Word Document, .docx
1033 Title of Data: Ion Proton sequencing primers.
1034 Description of Data: Oligo Name indicates the name of the forward or reverse primer; Oligo
1035 Names that contain numbers preceded by an "R" are reverse primers. Full Sequence notes the
1036 full Oligo sequence in the 5'-3' direction. Barcode sequence shows the unique genetic barcode
1037 within each primer, in the 5'-3' direction, used in multiplexing and demultiplexing libraries.
1038

1039 **File Name: Additional file 4**

1040 File Format: Excel document, .xlsx

1041 Title of Data: Estimated fitness change in 250 generations evolution and significance of 92

1042 barcoded yeast strains.

1043 Description of Data: Summary of results for a fixed-0-intercept linear model of mean-corrected

1044 barcoded yeast initial fitness. Predictors are individual barcoded yeast strain ID's. FDR adjusted

1045 p-values for a 2-tailed test are reported.

1046

1047 **File Name: Additional file 5**

1048 File Format: html document, .html

1049 Title of Data: Data table of power calculations using error in the POC fitness assays data to

1050 support figure 3 panel A.

1051 Description of Data: Data used to generate figure 3 panel A. Columns, in order, are: fitness

1052 change (fitchange), population standard deviation in the POC fitness assays data among

1053 replicate measures of fitness (psd), number of replicates per timepoint (n), power, and

1054 significance level (sig.level). fitchange and n are varied to calculate power; psd and sig.level are

1055 kept constant.

1056

1057 **File Name: Additional file 6**

1058 File Format: html document, .html

1059 Title of Data: Data table of power calculations using error in the POC fitness assays data to

1060 support figure 3 panel B.

1061 Description of Data: Data used to generate figure 3 panel B. Columns, in order, are: degrees of
1062 freedom in the numerator (u), degrees of freedom in the denominator (v), power, effect size
1063 (f2), and significance level (sig.level). v and f2 are varied to calculate power; u and sig.level are
1064 kept constant.
1065
1066 **File Name: Additional file 7**
1067 File Format: pdf document, .pdf
1068 Title of Data: barcode cross-contamination rates for sequenced libraries.
1069 Description of Data: A. Histogram of mean barcode contamination rate (as a percentage of total
1070 count) for the cross-contamination diagnostic samples included in the Proof of Concept Fitness
1071 Assays sequenced library. One datapoint is reported for each unique forward-reverse index pair
1072 (sample) in the sequenced library. B. Histogram of mean barcode contamination rate (as a
1073 percentage of total count) for samples included in the 250-generation experimental evolution
1074 project. For resequenced samples, only the less contaminated sample is retained. One
1075 datapoint is reported for each unique forward-reverse index pair (sample) in this consensus
1076 library.
1077
1078 **File Name: Additional file 8**
1079 File Format: pdf document, .pdf
1080 Title of Data: Decreased contamination after reprocessing samples.
1081 Description of Data: Violin plots showing decrease in barcode contamination rate for a set of
1082 samples from the 250-generation evolution experiment that were DNA extracted, PCR

amplified, and sequenced two separate times. Colors depict sequencing runs: cyan for run 1., and orange for run 2. Point sizes reflect the number of reads underlying each datapoint. Mean contamination rate for run 1., and run 2., are depicted as heavy black crossbars.

File Name: Additional file 9

File Format: html document, .html

Title of Data: Data table of power calculations using error in the change in fitness data for the 250-generation experiment to support figure 3 panel C.

Description of Data: Data used to generate figure 3 panel C. Columns, in order, are: fitness change (fitchange), population standard deviation in the POC fitness assays data among replicate measures of fitness (psd), number of replicates per timepoint (n), power, and significance level (sig.level). fitchange and n are varied to calculate power; psd and sig.level are kept constant.

File Name: Additional file 10

File Format: html document, .html

Title of Data: Data table of power calculations using error in the change in fitness data for the 250-generation experiment to support figure 3 panel D.

Description of Data: Data used to generate figure 3 panel D. Columns, in order, are: degrees of freedom in the numerator (u), degrees of freedom in the denominator (v), power, effect size (f2), and significance level (sig.level). v and f2 are varied to calculate power; u and sig.level are kept constant.

1105

1106 **File Name: Additional file 11**

1107 File Format: html document, .html

1108 Title of Data: Change in fitness over 250 generations of experimental evolution.

1109 Description of Data: Summary of results for a fixed-0-intercept linear model of barcoded yeast

1110 fitness change over 250-generations of experimental evolution. Predictors are generation-250

1111 barcoded yeast strains. FDR adjusted p-values for a 1-tailed (greater) test are reported.

1112

1113 **File Name: Additional file 12**

1114 File Format: Excel document, .xlsx

1115 Title of Data: Standard error in change in fitness across siz treatments in 250 generations of

1116 experimental evolution.

1117 Description of data: summary of results for linear model of standard error in change in

1118 barcoded yeast strain fitness over 250-generations of experimental evolution.

1119

1120 **File Name: Additional file 13**

1121 File Format: html document, .html

1122 Title of Data: Change in fitness across six treatments in 250 generations of experimental

1123 evolution.

1124 Description of Data: summary of results for linear mixed effects model of change in barcoded

1125 yeast strain fitness over 250-generations of experimental evolution.

1126

File Name: Additional file 14

File Format: html document, .html

Title of Data: t-max by treatment model results.

Description of Data: summary of results for linear mixed effects model of generation of maximum deviation in barcode abundance from initial conditions over 250-generations of experimental evolution by treatment.

File Name: Additional file 15

File Format: pdf document, .pdf

Title of Data: t-max by treatment visualization.

Description of Data: Violin plot of generation of maximum deviation in barcode abundance from initial conditions for 152 yeast strains evolved across six evolutionary treatments for 250 generations. Point sizes reflect the number of reads underlying each datapoint and colors indicate evolutionary treatments. Treatment means are depicted as heavy black crossbars. Treatments significantly different from the control treatment are marked with an asterisk. The treatment with diploid yeast evolved under a standard 1:1000 transfer dilution in CM is selected as the reference level in this model.

File Name: Additional file 16

File Format: html document, .html

Title of Data: m-max by treatment model results.

1148 Description of Data: summary of results for linear mixed effects model of maximum deviation in
1149 barcode abundance from initial conditions over 250-generations of experimental evolution by
1150 treatment.
1151
1152 **File Name: Additional file 17**
1153 File Format: pdf document, .pdf
1154 Title of Data: m-max by treatment visualization.
1155 Description of Data: Violin plot of magnitude of maximum deviation in barcode abundance
1156 from initial conditions for 152 yeast strains evolved across six evolutionary treatments for 250
1157 generations. Point sizes reflect the number of reads underlying each datapoint and colors
1158 indicate evolutionary treatments. Treatment means are depicted as heavy black crossbars.
1159 Treatments significantly different from the control treatment are marked with an asterisk. The
1160 treatment with diploid yeast evolved under a standard 1:1000 transfer dilution in CM is
1161 selected as the reference level in this model.
1162
1163 **File Name: Additional file 18**
1164 File Format: html document, .html
1165 Title of Data: t-max-rate by treatment model results.
1166 Description of Data: summary of results for linear mixed effects model of generation of
1167 maximum rate of change in barcode abundance over 250-generations of experimental
1168 evolution by treatment.
1169

1170 **File Name: Additional file 19**

1171 File Format: pdf document, .pdf

1172 Title of Data: t-max-rate by treatment visualization.

1173 Description of Data: Violin plot of generation of maximum rate of change in barcode abundance

1174 for 152 yeast strains evolved across six evolutionary treatments for 250 generations. Point sizes

1175 reflect the number of reads underlying each datapoint and colors indicate evolutionary

1176 treatments. Treatment means are depicted as heavy black crossbars. Treatments significantly

1177 different from the control treatment are marked with an asterisk. The treatment with diploid

1178 yeast evolved under a standard 1:1000 transfer dilution in CM is selected as the reference level

1179 in this model.

1180

1181 **File Name: Additional file 20**

1182 File Format: html document, .html

1183 Title of Data: m-max-rate by treatment model results.

1184 Description of Data: Summary of results for linear mixed effects model of magnitude of

1185 maximum rate of change in barcode abundance over 250-generations of experimental

1186 evolution by treatment.

1187

1188 **File Name: Additional file 21**

1189 File Format: pdf document, .pdf

1190 Title of Data: m-max-rate by treatment visualization.

Description of Data: Violin plot of magnitude of maximum rate of change in barcode abundance for 152 yeast strains evolved across six evolutionary treatments for 250 generations. Point sizes reflect the number of reads underlying each datapoint and colors indicate evolutionary treatments. Treatment means are depicted as heavy black crossbars. Treatments significantly different from the control treatment are marked with an asterisk. The treatment with diploid yeast evolved under a standard 1:1000 transfer dilution in CM is selected as the reference level in this model.

File Name: Additional file 22

File Format: html document, .html

Title of Data: t-max-diff by treatment model results.

Description of Data: Summary of results for linear mixed effects model of generation of maximum difference in sympatric barcode abundance over 250-generations of experimental evolution by treatment.

File Name: Additional file 23

File Format: pdf document, .pdf

Title of Data: t-max-diff by treatment visualization.

Description of Data: Violin plot of generation of maximum difference in sympatric barcode abundance for 152 yeast strains evolved across six evolutionary treatments for 250 generations. Point sizes reflect the number of reads underlying each datapoint and colors indicate evolutionary treatments. Treatment means are depicted as heavy black crossbars.

1213 Treatments significantly different from the control treatment are marked with an asterisk. The
1214 treatment with diploid yeast evolved under a standard 1:1000 transfer dilution in CM is
1215 selected as the reference level in this model.
1216
1217 **File Name: Additional file 24**
1218 File Format: html document, .html
1219 Title of Data: m-max-diff by treatment model results.
1220 Description of Data: Summary of results for linear mixed effects model of magnitude of
1221 maximum difference in sympatric barcode abundance over 250-generations of experimental
1222 evolution by treatment.
1223
1224 **File Name: Additional file 25**
1225 File Format: pdf document, .pdf
1226 Title of Data: m-max-diff by treatment visualization.
1227 Description of Data: Violin plot of magnitude of maximum difference in sympatric barcode
1228 abundance for 152 yeast strains evolved across six evolutionary treatments for 250
1229 generations. Point sizes reflect the number of reads underlying each datapoint and colors
1230 indicate evolutionary treatments. Treatment means are depicted as heavy black crossbars.
1231 Treatments significantly different from the control treatment are marked with an asterisk. The
1232 treatment with diploid yeast evolved under a standard 1:1000 transfer dilution in CM is
1233 selected as the reference level in this model.
1234

1235 **File Name: Additional file 26**

1236 File Format: html document, .html

1237 Title of Data: Total change in barcode abundance by treatment model results.

1238 Description of Data: summary of results for linear mixed effects model of total (cumulative)

1239 change in sympatric barcode abundance over 250-generations of experimental evolution by

1240 treatment.

1241

1242 **File Name: Additional file 27**

1243 File Format: pdf document, .pdf

1244 Title of Data: Total change in barcode abundance by treatment visualization.

1245 Description of Data: Violin plot of total (cumulative) change in sympatric barcode abundance

1246 for 152 yeast strains evolved across six evolutionary treatments for 250 generations. Point sizes

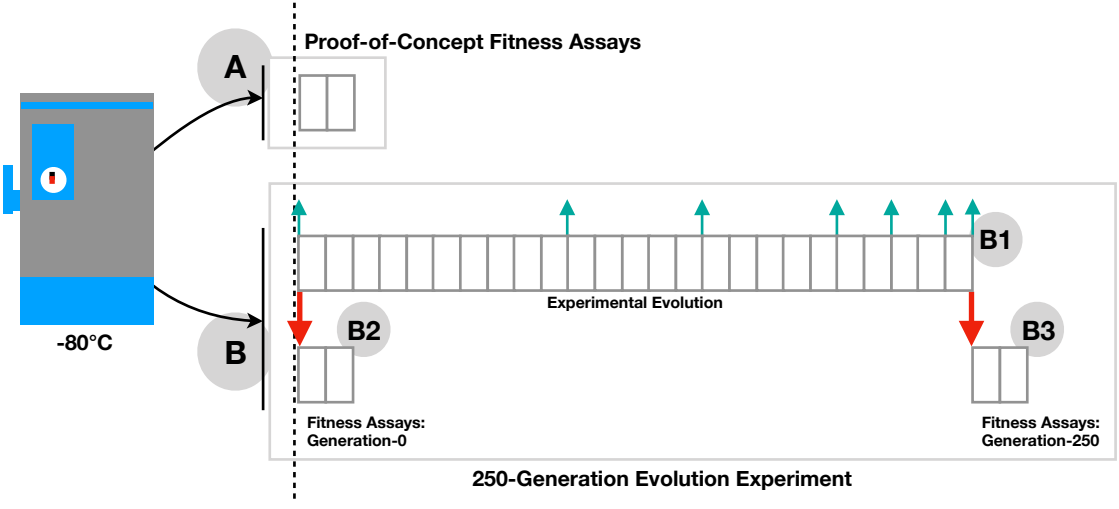
1247 reflect the number of reads underlying each datapoint and colors indicate evolutionary

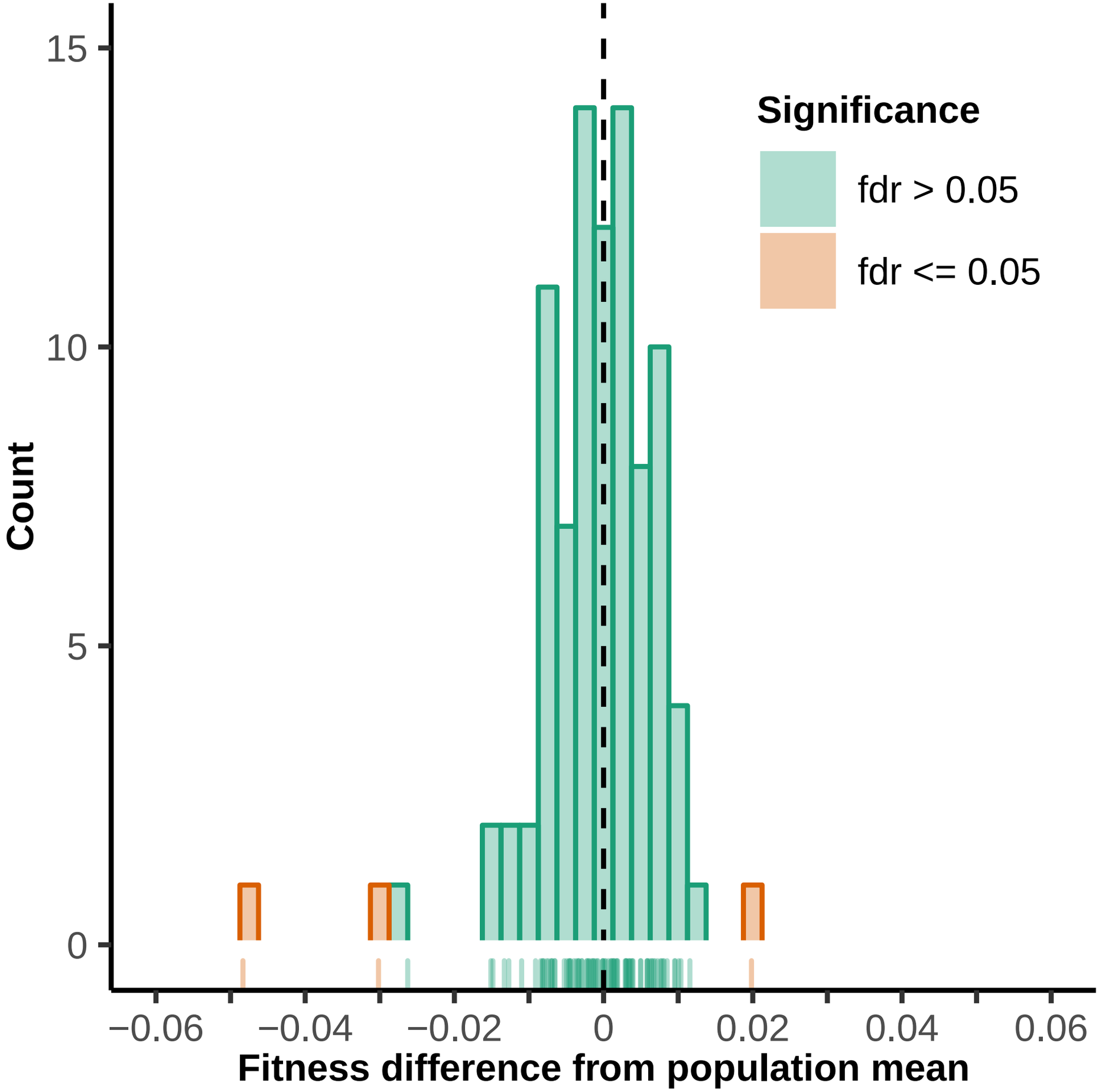
1248 treatments. Treatment means are depicted as heavy black crossbars. Treatments significantly

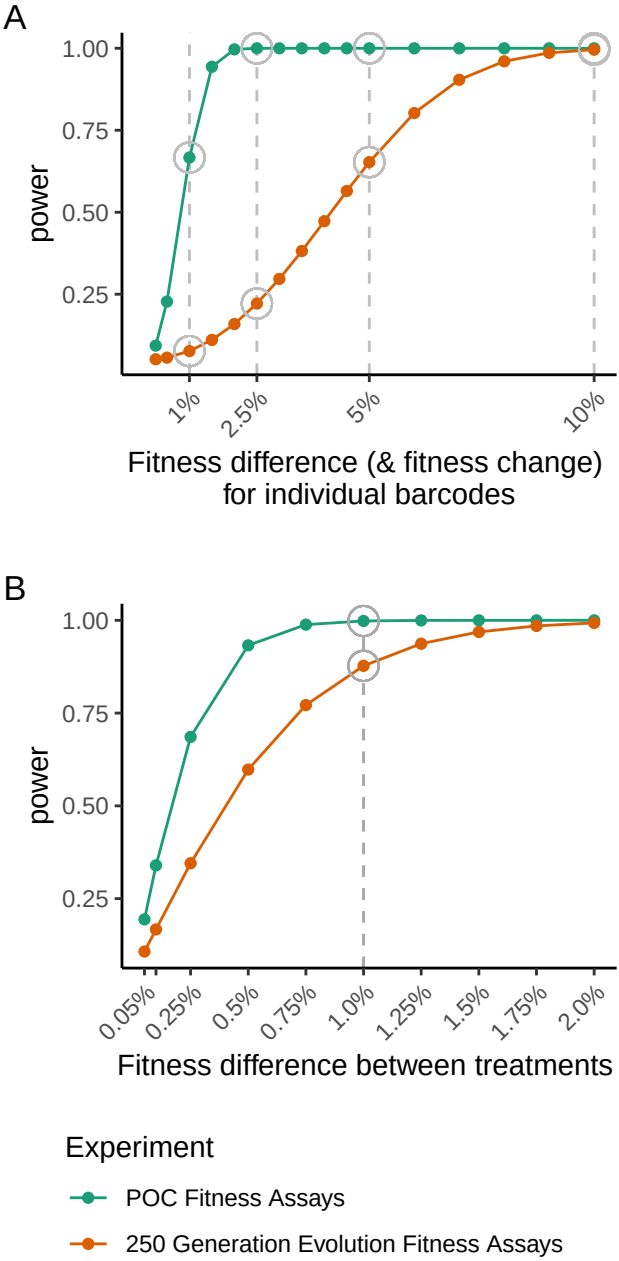
1249 different from the control treatment are marked with an asterisk. The treatment with diploid

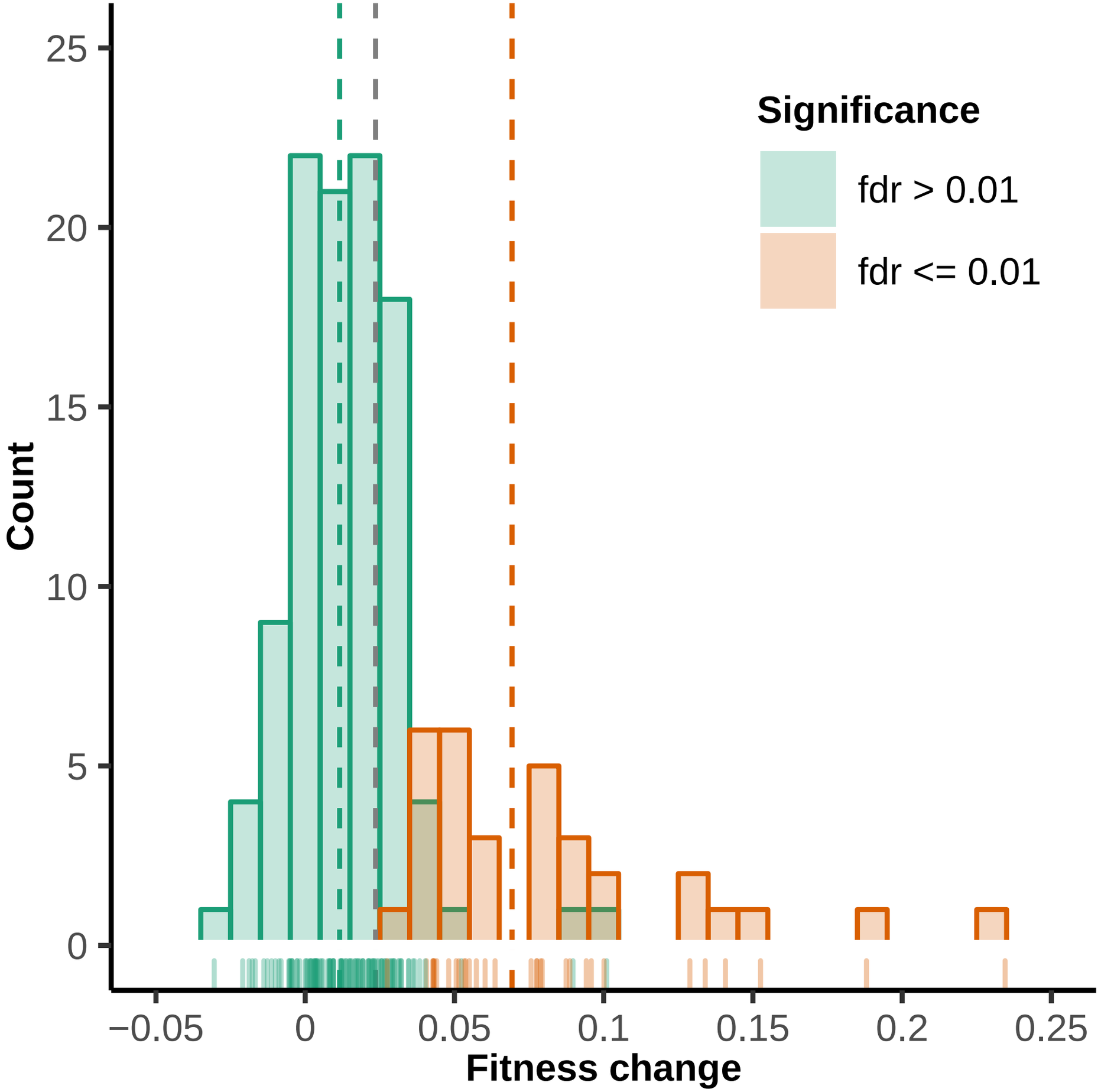
1250 yeast evolved under a standard 1:1000 transfer dilution in CM is selected as the reference level

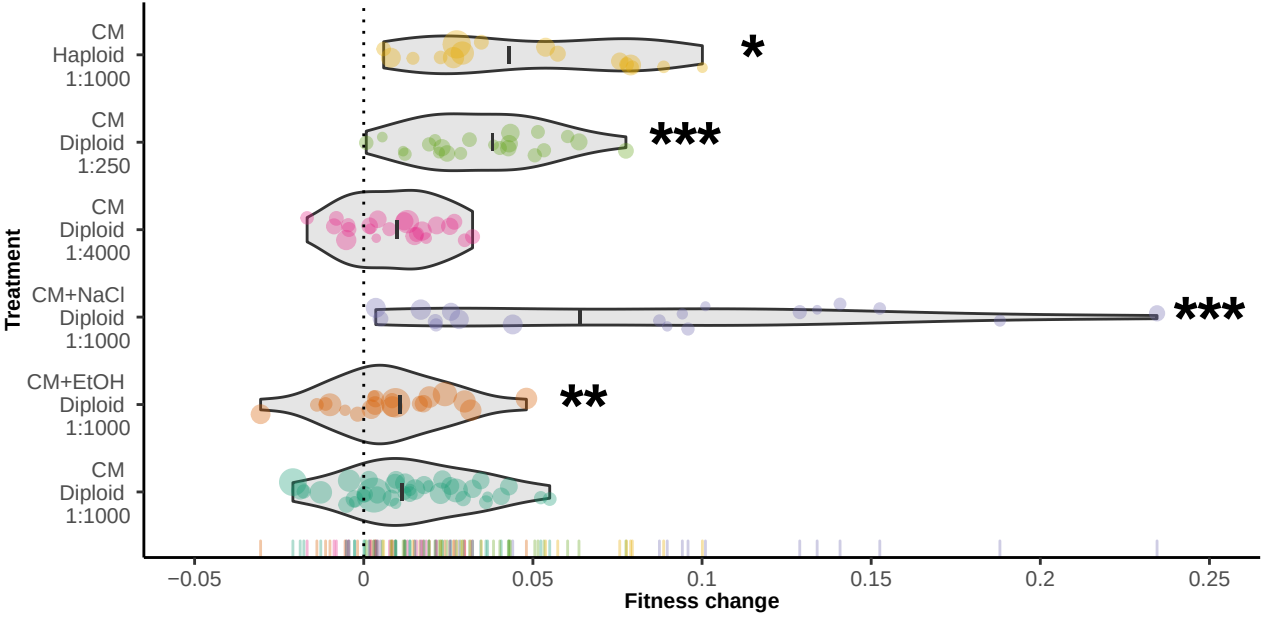
1251 in this model.


Proof-of-Concept Fitness Assays
A
B
-80°C
B1
B2
B3
Experimental Evolution
Fitness Assays:
Generation-0
Fitness Assays:
Generation-250
250-Generation Evolution Experiment


Significance
fdr > 0.05
fdr <= 0.05
Count
15
10
5
0
−0.06
−0.04
−0.02
0
0.02
0.04
0.06
Fitness difference from population mean


A
1.00
0.75
0.50
0.25
power
1%
2.5%
5%
10%
Fitness difference (& fitness change)
for individual barcodes
B
1.00
0.75
0.50
0.25
power
0.05%
0.25%
0.5%
0.75%
1.0%
1.25%
1.5%
1.75%
2.0%
Fitness difference between treatments
Experiment
POC Fitness Assays
250 Generation Evolution Fitness Assays


Significance
fdr > 0.01
fdr <= 0.01
Count
25
20
15
10
5
0
−0.05
0
0.05
0.1
0.15
0.2
0.25
Fitness change


CM
Haploid
1:1000
CM
Diploid
1:250
CM
Diploid
1:4000
CM+NaCl
Diploid
1:1000
CM+EtOH
Diploid
1:1000
CM
Diploid
1:1000
Treatment
*
***
***
**
−0.05
0
0.05
0.1
0.15
0.2
0.25
Fitness change