

1 **TITLE:**

2 A genotype-phenotype-fitness map reveals local modularity and global pleiotropy of adaptation

3

4 **AUTHORS:**

5 Grant Kinsler^{1*}, Kerry Geiler-Samerotte^{1,2*}, Dmitri Petrov¹

6

7 **AFFILIATIONS:**

8 ¹Department of Biology, Stanford University, Stanford, CA

9 ²Center of Mechanisms of Evolution, School of Life Sciences, Arizona State University, Tempe,
10 AZ

11 *Equal contribution

12

13 **SUMMARY:**

14 Building a genotype-phenotype-fitness map of adaptation is a central goal in evolutionary
15 biology. It is notoriously difficult even when the adaptive mutations are known because it is hard
16 to enumerate which phenotypes make these mutations adaptive. We address this problem by
17 first quantifying how the fitness of hundreds of adaptive yeast mutants responds to subtle
18 environmental shifts and then modeling the number of phenotypes they must collectively
19 influence by decomposing these patterns of fitness variation. We find that a small number of
20 phenotypes predicts fitness of the adaptive mutations near their original glucose-limited
21 evolution condition. Importantly, phenotypes that matter little to fitness at or near the evolution
22 condition can matter strongly in distant environments. This suggests that adaptive mutations are
23 locally modular—affecting a small number of phenotypes that matter to fitness in the
24 environment where they evolved—yet globally pleiotropic—affecting additional phenotypes that
25 may reduce or improve fitness in new environments.

26 INTRODUCTION

27
28 High-replicate laboratory evolution experiments are opening an unprecedented window into the
29 dynamics and genetic basis of adaptive change by *de novo* mutation (Croizat et al., 2010; Good
30 et al., 2017; Huang et al., 2018; Lang et al., 2013; Levy et al., 2015; Tenaillon et al., 2012;
31 Venkataram et al., 2016). One of the key insights revealed by these studies is that in many
32 systems, evolution can initially proceed rapidly via many large-effect single mutations. While the
33 identities of these adaptive mutations are often unique to a specific replicate of the evolutionary
34 experiment, across many replicates they tend to occur in similar functional units (e.g. genes and
35 pathways) (Croizat et al., 2010; Fumasoni and Murray, 2020; Good et al., 2017; Huang et al.,
36 2018; Lang et al., 2013; Levy et al., 2015; Tenaillon et al., 2012; Venkataram et al., 2019, 2016).
37 Thus, while the diversity of mutations suggests that there might be many ways to adapt, the
38 much smaller number of apparent functional units implies, in contrast, that most adaptive
39 mutations affect a small set of key phenotypes (Fig 1A).

40 Consider the seminal study by Tenaillon et al. (Tenaillon et al., 2012) in which 115 populations
41 were evolved at high temperature for ~2000 generations. While the authors identified over a
42 thousand mutations that were largely unique to each population, the number of affected genes
43 was much smaller with 12 genes being hit over 25 times each. Even greater convergence was
44 seen at higher levels of organization such as operons. Similarly, Venkataram et al (Venkataram
45 et al., 2016) found that, of the hundreds of unique genetic mutations that occur during
46 adaptation to glucose-limitation, the vast majority fall into a relatively small number of genes
47 (mostly *IRA1*, *IRA2*, *GPB2*, *PDE2*) and primarily two pathways - Ras/PKA and TOR/Sch9. Thus
48 despite the diversity of mutations, it is possible that all of their effects can be mapped in one or
49 few dimensions required to describe their effects on the Ras/PKA or TOR/Sch9 pathways.
50 These are just two examples, but the pattern has been seen repeatedly (Barghi et al., 2019;
51 Croizat et al., 2010; Good et al., 2017; Lang et al., 2013; Lind et al., 2015). Note that this pattern
52 is seen not only in experimental evolution but also in cancer evolution. Individual tumors are
53 largely unique in terms of specific mutations, but these mutations affect a much smaller set of
54 driver genes and an even smaller number of higher functional units such as signalling pathways
55 (Bailey et al., 2018; Hanahan and Weinberg, 2011, 2000; Sanchez-Vega et al., 2018; Sondka et
56 al., 2018).

57 The mapping of adaptive mutations to a smaller number of functional units and thus a low-
58 dimensional space representing the small number of phenotypes that they collectively affect
59 (Fig 1A) is consistent with theoretical models of adaptation. These theoretical models argue that
60 adaptive mutations, especially those of substantial fitness benefit, cannot affect too many
61 phenotypes at once as most such effects should be deleterious and thus inconsistent with the
62 overall positive effect on fitness (Fisher, 1930; Orr, 2000). More recent studies likewise suggest
63 that selection against mutations with high pleiotropy, *i.e.* mutations that affect many phenotypes,
64 has resulted in a modular architecture of the genotype-phenotype map, in which genetic
65 changes can influence some phenotypes without affecting others (Altenberg, 2005; Collet et al.,
66 2018; Hartwell et al., 1999; Melo et al., 2016; Wagner et al., 2007; Wagner and Altenberg, 1996;
67 Wagner and Zhang, 2011; Welch and Waxman, 2003). This architecture would allow single
68 mutations to have a large effect on a small number of important phenotypes. It would also
69 explain the observations that very large collections of adaptive mutations are not diverse in
70 terms of affected genes, pathways, and phenotypes. The reason for this is that only mutations
71 that affect the genes, pathways, and phenotypes that represent the correct module most
72 relevant to adaptation in the specific environment will be adaptive.

73 While theoretically appealing, the possibility that observed adaptive mutations indeed affect only
74 a very small number of phenotypes is difficult to reconcile with the notion that organisms are
75 tightly integrated (Kacser and Burns, 1981; Paaby and Rockman, 2013; Rockman, 2012).

76 Further, there is experimental evidence of widespread pleiotropy, for example, from genome
77 wide association studies that suggest every gene can influence every trait, at least to some
78 extent (Boyle et al., 2017; Chesmore et al., 2018; Sella and Barton, 2019; Sivakumaran et al.,
79 2011; Visscher and Yang, 2016). It is possible that pleiotropy is common, but strongly adaptive
80 mutations observed in experimental evolution are unusual in that they have few pleiotropic
81 phenotypic effects. Another possibility is that these mutations do have pleiotropic side effects
82 but these matter little to fitness in the evolution condition (Fig 1B, left side). Note that here we do
83 not need to claim that these phenotypic effects *never* matter to fitness but rather that they do not
84 matter substantially to fitness in the condition where they evolved. In fact the key prediction of
85 this model is that one should be able to detect such latent pleiotropy by showing that additional
86 phenotypic changes matter to fitness in other environments (Fig 1B, right side).

87 If the model depicted in Fig 1B is true, then it is possible that adaptive mutations are *locally*
88 *modular* — that they affect very few phenotypes that matter to fitness in the evolution condition
89 — and *globally pleiotropic*. Under this model, the large number of distinct mutations available to
90 adaptation becomes important. Indeed while these mutations tend to influence similar genes
91 and pathways, their phenotypic effects do not simply collapse to a low dimensional space.
92 Instead this genetic diversity becomes a source of consequential phenotypic diversity, but only
93 once these genetic variants leave the local environment in which they originated.

94 In order to test this model and better understand the genotype-phenotype-fitness map, we face
95 the difficult task of identifying which phenotypes are affected by the adaptive mutations and then
96 determining how these phenotypes contribute to fitness. This is a challenging problem as the
97 possible number of phenotypes one can measure is effectively infinite, e.g. the expression level
98 of every gene or the quantity of every metabolite (Coombes et al., 2019; Mehlhoff et al., 2020).
99 Further, many measurable phenotypes are related in complex ways (Geiler-Samerotte et al.,
100 2019). Mapping their contribution to fitness requires a complete understanding of how genetic
101 changes lead to molecular changes and how these percolate to higher functional levels and
102 ultimately influence fitness (Kemble et al., 2020). This might be possible to do in some cases
103 where the phenotype to fitness mapping is simple (e.g. antibiotic resistance driven by a specific
104 enzyme or tRNA or protein folding mediating specific RNA or protein function) (Baeza-Centurion
105 et al., 2019; Cowperthwaite et al., 2005; Diss and Lehner, 2018; Domingo et al., 2019; Harmand
106 et al., 2017; Karageorgi et al., 2019; Li and Zhang, 2018; Otwinowski et al., 2018; Pressman et
107 al., 2019; Sarkisyan et al., 2016; Starr et al., 2018; Weinreich, 2006) but is exceptionally difficult
108 for complex phenotypes.

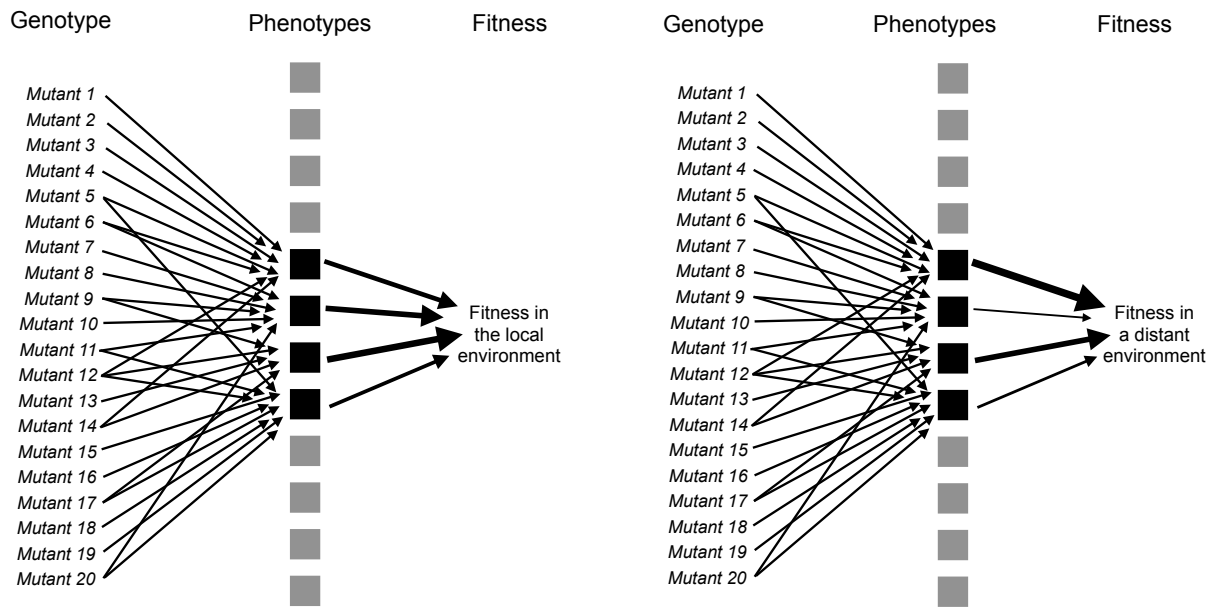
109 Moreover, to distinguish between the model in which mutations affect a small number of
110 phenotypes (Fig. 1A) and the model in which mutations affect many phenotypes, albeit with few
111 contributing substantially to fitness in the evolution condition (Fig 1B), we need to understand
112 these genotype-phenotype-fitness maps not only in the environment in which adaptive mutants
113 evolved, but also in other environments. And we need to do this for many adaptive mutants so
114 that we can assess the extent to which different mutants affect different phenotypes.
115 Considering the scope of this challenge, it is not surprising that despite much theoretical
116 discussion of modularity and pleiotropy as it relates to adaptation, experimental approaches to
117 address these questions have lagged behind.

118 Here we suggest a way to model the genotype-phenotype-fitness relationship that avoids the
119 problem of measuring each phenotype and its effect on fitness explicitly. We argue that it is
120 possible to investigate the genotype-phenotype-fitness map by comparing how the fitness
121 effects of many mutations change across a large number of environments. The way each
122 mutant's fitness varies across environments must be related to its phenotype, and thus the way
123 mutants co-vary in fitness across environments tells us whether they affect similar fitness-
124 relevant phenotypes. We can use the profiles of fitness across a set of environments to identify

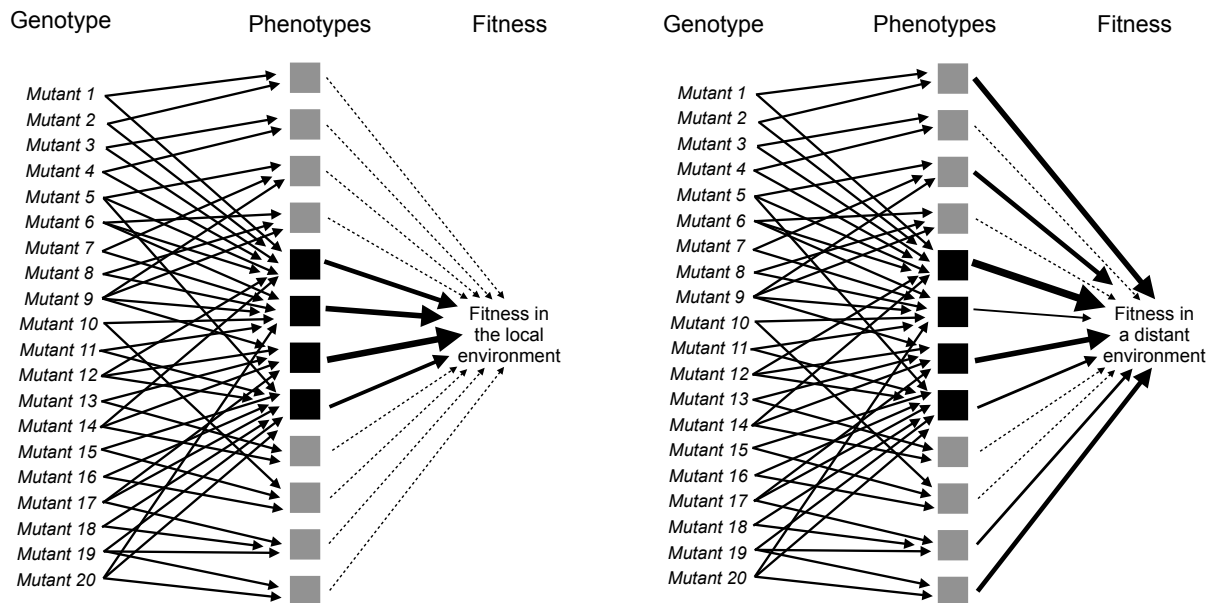
125 the total number of fitness-relevant phenotypes affected across a collection of adaptive mutants,
126 the extent to which different mutants affect different phenotypes, and whether the contribution of
127 each phenotype to fitness changes across environments.

128
129 Here we build a genotype-phenotype-fitness model for hundreds of adaptive yeast mutants that
130 originally evolved in a glucose-limited environment. We use this model to accurately predict the
131 fitness of these mutants across a set of 45 environments that vary in their similarity to the
132 evolution condition. We find that the behavior of adaptive mutations can be described by a low-
133 dimensional phenotypic model. In other words, these mutants affect a small number of
134 phenotypes that matter to fitness in the glucose-limited condition in which they evolved. We find
135 that this low-dimensional phenotypic model makes accurate predictions of mutant fitness in
136 novel environments even when they are distant from the evolution condition. Moreover, we find
137 that some phenotypes that contribute very little to fitness in the evolution condition become
138 surprisingly important in some novel environments. This suggests that adaptive mutations are
139 globally pleiotropic in that they affect many phenotypes overall, but that they are locally modular
140 in that only a small number of these phenotypes have substantial effects on fitness in the
141 environment they evolved in. Overall, we suggest that this set of adaptive mutations contains
142 substantial and consequential latent phenotypic diversity, meaning that despite targeting similar
143 genes and pathways, different adaptive mutants may respond differently to future evolutionary
144 challenges. This finding has important consequences for understanding how directional
145 selection can generate consequential phenotypic heterogeneity both in natural populations and
146 also in the context of diseases such as cancer and viral or bacterial infections. In addition, our
147 results show that our abstract, top-down approach is a promising route of analysis for
148 investigating the phenotypic and fitness consequences of mutation.

A



B



149 **Figure 1. Adaptive mutations can be locally modular and globally pleiotropic.** (A) A collection of
 150 adaptive mutations may affect a small number of phenotypes (four black squares). If these adaptive
 151 mutations only affect these phenotypes, then fitness in both the environment they evolved in (local
 152 environment) and other environments (distant environment) is determined solely by these phenotypes.
 153 (B) Alternatively, these mutations may collectively (and individually) affect many phenotypes, but only a
 154 small number of phenotypes may matter to fitness (those indicated by black squares with thick arrows
 155 pointing to fitness), whereas the other phenotypes may make very small contributions to fitness (those
 156 indicated by the gray squares and thin, dashed lines leading to fitness). Under this model, the contribution
 157 of each phenotype to fitness can change depending on the environment. Thus fitness differences
 158 between seemingly similar mutants can be revealed by measuring fitness in more environments. Such
 159 fitness differences suggest the presence of phenotypic differences between mutants.

160
 161 **RESULTS**

162 **Mutants that improve fitness under glucose limitation vary in their genotype-by-**
 163 **environment interactions**

164

165 A previous evolution experiment generated a collection of hundreds of adaptive yeast mutants,
166 each of which typically harbors a single independent mutation that provides a benefit to growth
167 in a glucose-limited environment (Levy et al., 2015). Many of these mutants, which began the
168 evolution experiment as haploids, underwent whole-genome duplication to become diploid,
169 which improved their relative fitness (Venkataram et al 2016). Some of these diploids acquired
170 additional mutations, including increased copy number of either chromosome 11 or 12 as well
171 as point mutations, which generated additional fitness benefits. The adaptive mutants that
172 remained haploid acquired both gain- and loss-of-function mutations in nutrient-response
173 pathways (Ras/PKA and TOR/Sch9). Some other mutations were also observed, including a
174 mutation in the HOG pathway gene *SSK2* (Venkataram et al., 2016). Although these mutants
175 have been well-characterized at the level of genotype and fitness, it is unclear what phenotypes
176 they affect. The first question we address is whether these diverse mutations collectively affect
177 a large number of phenotypes that matter to fitness, or whether these mutants are functionally
178 similar in that they collectively alter a small set of fitness-relevant phenotypes.

179
180 Understanding the map from genotype to phenotype to fitness is extremely challenging because
181 each genetic change can influence multiple traits, not all of which are independent or contribute
182 to fitness in a meaningful way. We contend with this challenge by measuring how the relative
183 fitness of each adaptive mutant changes across a large collection of similar and dissimilar
184 environments, which we term the “fitness profile”. When a group of mutants demonstrate similar
185 responses to environmental change, we conclude that these mutants affect similar phenotypes.
186 By clustering mutants with similar fitness profiles across a collection of environments, we can
187 learn about which mutants influence similar phenotypes, as well as estimate the total number of
188 fitness-relevant phenotypes represented across all mutants and all investigated environments.

189
190 Because our mutant strains are barcoded, we can use previously-established methods to
191 measure their relative fitness in bulk and with high-precision (Venkataram et al., 2016).
192 Specifically, we compete a pool of the barcoded mutants against an ancestral reference strain
193 over the course of several serial dilution cycles. During each 48 hour cycle, the yeast are given
194 fresh glucose-limited media which supports 8 generations of exponential growth after which
195 glucose is depleted and cells transition to non-fermentable carbon sources. After every 48 hour
196 cycle, we transfer $\sim 5 \times 10^7$ cells to fresh media to continue the growth competition. We also
197 extract DNA from the remaining cells to PCR amplify and sequence their barcodes. We repeat
198 this process four times, giving us an estimate of the frequency of each barcode at five time-
199 points. By quantifying the log-linear changes in each barcode’s frequency over time and
200 correcting for the mean-fitness change of population, we can calculate the fitness of each
201 barcoded mutant relative to the reference strain (Fig 2A; Methods).

202
203 Using this method, we quantify the fitness of a large number of adaptive mutants in 45
204 environments. We focus on a set of 292 adaptive mutants that have been sequenced, show
205 clear adaptive effects in the glucose-limited condition in which these mutants evolved (hereafter
206 “evolution condition”; EC) (Fig 2B; Table S1), and for which we obtained high-precision fitness
207 measurements in all 45 environments. These environments include some experiments from
208 previously published work (Li et al., 2018; Venkataram et al., 2016), as well as 32 new
209 environments including replicates of the evolution condition, subtle shifts to the amount of
210 glucose, changes to the shape of the culturing flask, changes to the carbon source, and addition
211 of stressors such as drugs or high salt (Table S2).

212
213 In order to determine the total number of phenotypes that are relevant to fitness in the EC, we
214 focus on environments that are very similar to the EC but still induce small yet detectable
215 perturbations in fitness. We do so because the phenotypes that are the most relevant to fitness
216 may change with the environment (Fig 1B). Thus, we partition the 45 environments into a set of
217 “subtle” perturbations, from which we will detect the phenotypes relevant to fitness near the EC,

218 and “strong” perturbations which we will use to study whether these mutants influence additional
219 phenotypes that matter in other environments (Fig 1B).

220

221 To partition environments into subtle and strong perturbations of the EC, we rely on the nested
222 structure of replicate experiments performed in the EC. We performed nine such replicates,
223 each at different times, which each included multiple replicates performed at the same time. We
224 observe much less variation across replicates performed simultaneously than across replicates
225 performed at different times ($p < 1e-5$ from permutation test). Variation across experiments
226 performed at different times is often referred to as “batch effects” and likely reflects
227 environmental variability that we were unable to control (e.g. slight fluctuations in incubation
228 temperature due to limits on the precision of the instrument). These environmental differences
229 between batches are very subtle, as they represent the limit of our ability to minimize
230 environmental variation. Thus, variation in fitness across the EC batches serves as a natural
231 benchmark for the strength of environmental perturbations. If the deviations in fitness caused by
232 an environmental perturbation are substantially stronger than those observed across the EC
233 batches, we call that perturbation “strong”.

234

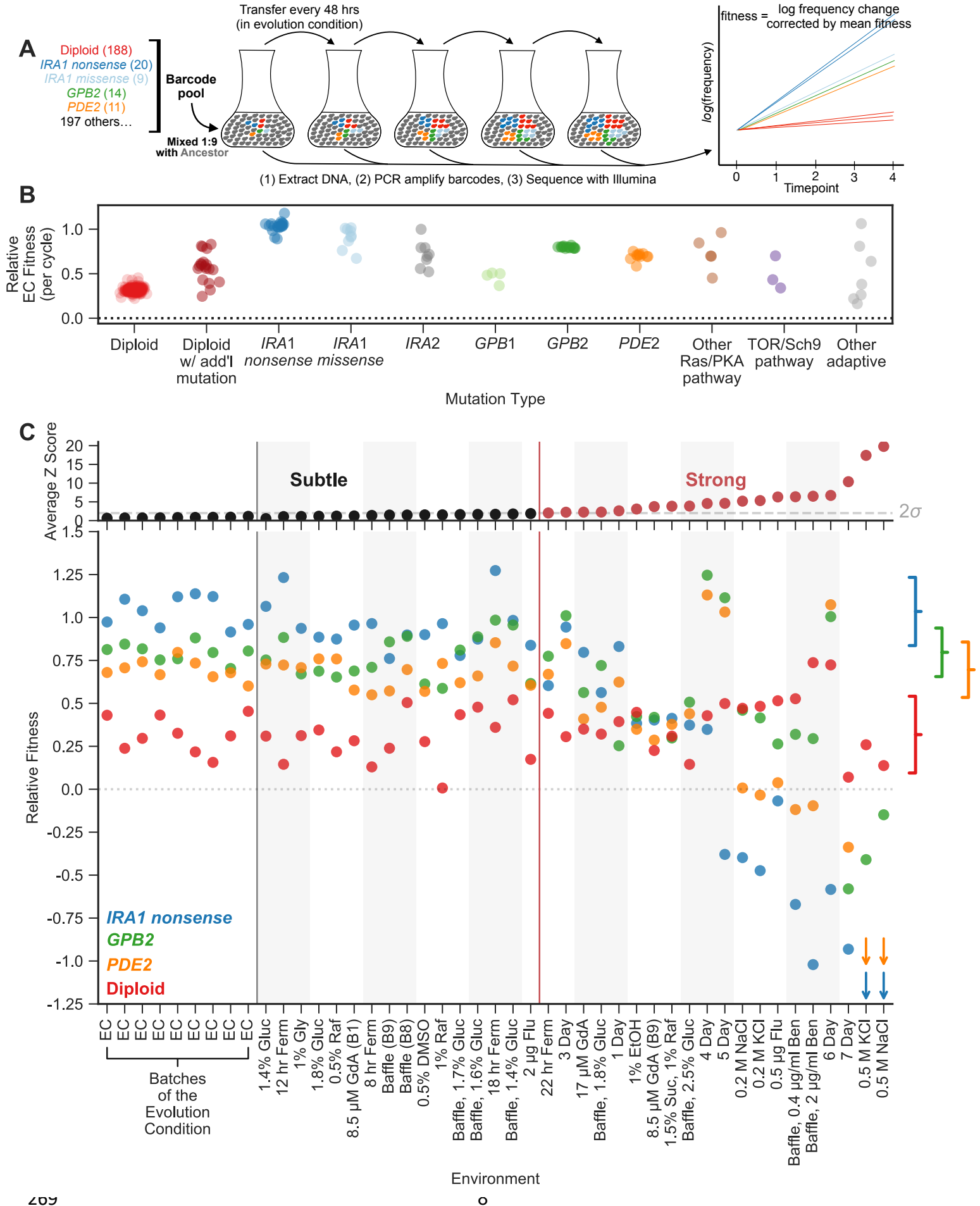
235 More explicitly, to determine whether a given environmental perturbation is subtle or strong, we
236 subtract the fitness of adaptive mutants in this environment from their average across the EC
237 batches. We then compare this difference to the variation in fitness observed across the EC
238 batches. Sixteen environmental perturbations provoked fitness differences that were similar to
239 those observed across EC batches (Z-score < 2). These environments, together with the nine
240 EC batches, make up a set of subtle environmental perturbations. The remaining 20
241 environments, where the average deviation in fitness is substantially larger than that observed
242 across batches (Z-score > 2), were classified as strong environmental perturbations (Fig 2C,
243 top; Methods).

244

245 The rank order of the fitnesses of many mutations is largely preserved across the 25
246 environments that represent subtle perturbations (Fig 2C, bottom). For example, *IRA1* nonsense
247 mutants, which are the most adaptive in the EC, generally remain the most adaptive across the
248 subtle perturbations. Additionally, the *GPB2* and *PDE2* mutants have similar fitness effects
249 across EC batches and only occasionally switch order across the subtle environmental
250 perturbations. In contrast, the 20 environments that represent strong perturbations reveal clear
251 genotype-by-environment interactions (Fig 2C, bottom). For example, altering the transfer time
252 from 48 to 24 hours (the “1 Day” environment in Fig 2C) affects *GPB2* mutants more strongly
253 compared to the other mutants in the Ras/PKA pathway, including *IRA1* and *PDE2*. The
254 strongest environmental perturbations reveal clear tradeoffs for some of these adaptive
255 mutants. For example, *PDE2* and *IRA1-nonsense* but not *GPB2* mutants are particularly
256 sensitive to osmotic stress as indicated by the NaCl and KCl environments. Additionally, *IRA1-*
257 *nonsense* mutants become strongly deleterious in the long transfer conditions that experience
258 stationary phase (5-, 6-, 7-Day environments) (Li et al., 2018). In contrast to complex behavior
259 exhibited by the adaptive haploids, the diploids appear to be relatively robust to strong tradeoffs,
260 appearing similarly adaptive across all perturbations, subtle and strong.

261

262 The observation that different mutants have different and fairly complex fitness profiles suggests
263 that they have different phenotypic effects. Even *PDE2* and *GPB2*, which have similar fitnesses
264 in the EC and are negative regulators of the same signalling pathway, have different fitness
265 profiles. Do these diverse phenotypic effects contribute to fitness in the EC? To examine how
266 many phenotypes matter to fitness in the EC, we test whether it is possible to create low
267 dimensional models that capture the complexity of the fitness profiles of all adaptive mutants
268 across all subtle perturbations.



270 **Figure 2. Measuring fitness for a collection of adaptive mutants across many environments**
271 **reveals gene-by-environment interactions. (A)** Schematic of fitness measurement procedure. Adaptive
272 mutants tagged with DNA barcodes are pooled at a 1:9 ratio with an ancestral reference strain. The pool
273 is then propagated for several growth cycles, where the population is diluted into fresh media at fixed time
274 intervals. DNA is extracted from each time-point, and the barcode region is PCR amplified and then
275 sequenced. A mutant's relative fitness is calculated based on the rate of change of its barcode's
276 frequency, corrected for the mean fitness of the population (see Methods). Relative fitness is calculated in
277 units of "per cycle", representing the improvement of each barcode relative to the reference over the
278 course of the time between transfers. **(B)** Relative fitness of each mutant in the evolution condition,
279 calculated as the average across all 9 Evolution Condition (EC) batches. **(C) (top)** Environments are
280 ordered from left to right depending on the degree to which they perturb mutant fitness from the average
281 fitness observed across all EC batches. Environments in which average mutant fitness is within two
282 standard deviations of average mutant fitness across EC batches are denoted in black and make up the
283 subtle perturbation set. Environments in which aggregate mutant behavior exceeds two standard
284 deviations are shown in red and make up the strong perturbations set. **(bottom)** This plot displays, for the
285 four most common types of adaptive mutation observed in response to glucose limitation (Venkataram et
286 al., 2016), the average fitness in each of the 45 environments we study. Brackets on the right represent
287 the amount of variation in fitness observed for each type of mutation across the EC batches, with the
288 notch representing the mean and the arms representing two standard deviations on either side of the
289 mean.

291 **A model including 8 fitness-relevant phenotypes captures fitness variation across subtle** 292 **environmental perturbations**

293
294 We utilize these complex fitness profiles to estimate the number of phenotypes that contribute to
295 fitness in the EC. Given that many of these mutants affect genes in the same nutrient response
296 pathway, the number of unique phenotypes they affect may be small. Alternatively, given the
297 observation that these mutants have different interactions with environments that represent
298 strong perturbations (Fig 2C), this number may be large. We use singular value decomposition
299 (SVD) to ask how much of the complexity in these fitness profiles can be captured by a low-
300 dimensional phenotypic model (Fig 3A). SVD is a dimensionality reduction approach which here
301 decomposes fitness profiles into two abstract multi-dimensional spaces described below.

302
303 The first space, P , represents the phenotypic effects of mutants, where each phenotype is
304 represented as a dimension (there are k phenotypic dimensions depicted in Fig 3A). Each
305 mutant is represented by coordinates specifying a location in the phenotype space P (e.g.
306 mutant 1 having coordinates $(p_{11}, p_{12}, p_{13}, \dots, p_{1k})$). The ancestral reference lineage, which, by
307 definition, has relative fitness zero in every environment, is placed at the origin (e.g. $(0, 0, 0, \dots$
308 $0)$) in this phenotypic space. In this sense, we can think of a mutation's effect on any phenotype
309 as a measure of the distance from the location of the mutant in that phenotypic dimension to the
310 origin.

311
312 The second space, E , represents the contribution of each of the phenotypes in P to fitness, and
313 thus has the same number of dimensions as P . If a phenotype does not contribute substantially
314 to fitness in any environment, it is not represented as a dimension in either space. Therefore,
315 our model captures only fitness-relevant phenotypes. In space E , each environment is
316 represented by coordinates specifying a location (e.g. environment 1 having coordinates
317 $(e_{11}, e_{22}, e_{33}, \dots, e_{k1})$). These coordinates in E reflect the contribution (weight) of each of the k
318 phenotypic dimensions on fitness in that environment. For example, an environment where only
319 a single phenotype matters to fitness would be placed at the origin for all the axes, except for
320 the axis corresponding to the single phenotypic dimension that matters. Environments for which
321 the same phenotypes contribute to fitness will be placed closer together in the space E .

322
323 In this model, each phenotype contributes to fitness independently, by definition, such that the
324 fitness of mutant i in environment j is determined by each phenotypic effect of mutant i , scaled

325 by the contribution of that phenotype to fitness in environment j . A linear combination of these
326 weighted phenotypic effects determines the fitness of mutant i in environment j :

$$327 \quad f_{ij} = p_{i1}e_{1j} + p_{i2}e_{2j} + p_{i3}e_{3j} + \dots + p_{ik}e_{kj}$$

328
329 In this model, mutants with similar fitness profiles, for example mutants 1 and 2 in Fig 3A, will be
330 inferred as having similar phenotypic effects, and thus be located near each other in the
331 phenotypic space P . Mutants with dissimilar fitness profiles, for example mutants 3 and 4 in Fig
332 3A, can be inferred to have at least some differing phenotypic effects, which might be mediated
333 by a different effect on a single phenotypic component or different effects on many. Mutants with
334 dissimilar fitness profiles are informative about the number of dimensions needed in this
335 abstract model of phenotypic space.

336
337 This genotype-phenotype-fitness model that we generate using SVD harkens to Fisher's
338 geometric model (FGM), which defines an abstract space of orthogonal phenotypes relevant to
339 fitness (Fisher, 1930). Others have utilized FGM to answer questions about the number of
340 phenotypes affected by mutations, though most previous work focuses on deleterious mutations
341 and how their impacts vary across genetic backgrounds rather than environments (Blanquart et
342 al., 2014; Blanquart and Bataillon, 2016; Lourenço et al., 2011; Martin and Lenormand, 2006;
343 Poon and Otto, 2000; Tenaillon et al., 2007; Weinreich and Knies, 2013). A key difference
344 between FGM and our model is that our model does not make assumptions about the
345 distribution of phenotypic effects or whether the relationship between mutations in phenotype
346 space is additive.

347
348 Here, we utilize SVD to count the number of phenotypes that contribute to fitness in the original
349 glucose-limited environment in which these adaptive mutants evolved. We used SVD to build an
350 abstract model that captures fitness profiles of all 292 adaptive mutants across the 25 subtle
351 perturbations. This model suggests that the majority of the variation in fitness for the 292
352 adaptive mutants across the 25 subtle perturbations can be explained by eight phenotypic
353 dimensions. The first phenotypic component is very large and explains 95% of variation in
354 fitness across all mutants and all subtle perturbations (Fig 3B). This component captures the
355 variation in fitness explainable in the absence of genotype-by-environment interactions, where
356 each mutation has a single effect that is scaled by the environment. As such, this first
357 component effectively represents each mutant's average fitness in the EC (Fig S2) and the
358 average impact of each subtle perturbation on mutant fitness (Fig S2). It is not surprising that
359 this component explains much of this variation, as the fitness of mutants in the EC should be
360 predictive of fitness in similar environments. The next seven components capture additional
361 variation not detectable from the simple 1-component model and thus represent genotype-by-
362 environment interactions. Of these, the first four capture 87% of the variation not captured by
363 component one (67.8%, 8.3%, 5.6%, and 5.3%, respectively). The remaining three interaction
364 components each capture less than 2% of the variation not captured by component one (Fig
365 3B). We cannot distinguish any additional components, beyond these eight, from noise. This is
366 because we see components that explain a similar amount of variation when we apply SVD to
367 datasets composed exclusively of values generated by our noise model (Fig3B; see Methods
368 and FigS1 for additional details).

369
370 We confirm that these eight phenotypic components capture meaningful biological variation in
371 fitness by using bi-cross-validation. Specifically, we designate a balanced set of 60 of the 292
372 mutants as a training set, chosen such that the recurrent mutation types — diploids, high-fitness
373 diploids, Ras/PKA mutants — are roughly equally represented (see Methods). The remaining
374 232 mutants comprise the test set. This set contains all mutation types represented by only a
375 single mutant, including all TOR/Sch9 (*TOR1*, *SCH9*, *KOG1*) and HOG (*SSK2*) pathway
376 representatives, as well as the rest of the recurrent mutants that were not picked for the training
377 set. We iteratively construct phenotype spaces using the 60 training mutants while holding out

378 one subtle perturbation at a time and creating the space with the data from the remaining 24
379 subtle perturbations. We then predict the fitness of the 232 held-out testing mutants in the held-
380 out condition. We do so using all 8 components, and again with only 7, 6, and so on. Then, we
381 ask whether the 8 component model does a better job at predicting mutant fitness than the
382 other, lower dimensional models. If a component reflects measurement noise rather than
383 biological signal, then the inclusion of this component would lead to overfitting and should harm
384 the model's ability to predict fitness in the held-out data. Instead we find that, on average across
385 the 25 iterations, prediction power improves from the inclusion of each of the eight components.
386 This confirms that even the smallest of these components captures biologically meaningful
387 variation in fitness across the 25 subtle perturbations of the EC. However, the gain in predictive
388 power decreases for each component. The model with only the first component explains on
389 average 85% of weighted variance for the test mutants in the left-out conditions. A model with
390 only the top five components explains 95.1%, and all eight components explain 96.2% of
391 variation. This suggests that the last few components have very small contributions to fitness in
392 the environments near the EC.

393

394 **A model including 8 fitness-relevant phenotypes recapitulates known features of** 395 **adaptive mutations**

396

397 We next ask whether the 8-dimensional phenotypic model clusters adaptive mutants found in
398 similar genes or pathways (e.g. Ras/PKA or TOR/Sch9), or that represent similar mutation types
399 (haploid v. diploid). Alternatively, our model may classify mutations into functional units (*i.e.*
400 mutations that have similar phenotypic effects) in a way that does not conform to gene or
401 pathway identity. We use Uniform Manifold Approximation and Projection (UMAP) to visualize
402 the distance between all the mutants in this phenotypic space. Since the first phenotypic
403 dimension captures the average fitness of each mutant in the EC, and since we already know
404 that mutations to the same gene have similar fitness in the EC (Fig. 2B), we exclude the first
405 phenotypic dimension from this analysis, though the inclusion of the first component does not
406 change the identity of the clusters (Fig S3). By focusing on the other 7 components, we are
407 asking whether genotype-by-environment interactions also cluster the mutants by gene,
408 mutation type, and pathway.

409

410 These 7 genotype-by-environment interactions indeed tend to cluster the adaptive mutants by
411 type and by gene (Fig 3C). Specifically, the diploids, *IRA1-nonsense*, *GPB2*, and *PDE2* mutants
412 each form distinct clusters ($p = 0.0001$, $p = 0.006$, $p = 0.0001$, and $p = 0.0001$, respectively). To
413 generate p-values, we calculated the median pairwise distance, finding that multiple mutations
414 in the same cluster are indeed more closely clustered than randomly chosen groups of mutants.
415 Interestingly, the three smallest components, which capture very little variation in fitness across
416 the environments that reflect subtle perturbations of the EC, also cluster some mutants by gene
417 (Fig S3). Specifically, *PDE2*, *GPB2*, and *IRA1-nonsense* mutants are each closer to mutants of
418 their own type than to other adaptive haploids ($p = 0.0001$, $p = 0.0001$, and $p = 0.03$,
419 respectively). Note that the space defined by the three smallest components does not cluster
420 *IRA1-nonsense* mutants away from diploids ($p = 0.718$). This suggests that some mutants, e.g.
421 *IRA1-nonsense* and diploids, have smaller effects on these three phenotypic components.
422 Overall, our abstract phenotypic model, which reflects the way that each mutant's fitness
423 changes across environments, reveals that mutations to the same gene tend to interact similarly
424 with the environment. This suggests that our approach, like others that compare genotype-by-
425 environment interactions (Li et al., 2018), is a useful and unbiased way to identify mutations that
426 share functional effects.

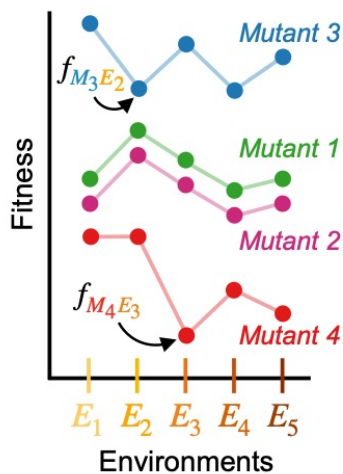
427

428 Our approach also detects cases where mutations to the same gene or pathway do not cluster
429 together. This suggests that our model captures phenotypic effects that would be obscured by
430 assuming mutations to the same gene affect the same traits. For example, genotype-by-

431 environment interactions do not cluster *IRA1-missense* mutations ($p = 0.317$) (Fig 3C; light blue
 432 points), despite clustering the *IRA1-nonsense* mutations. Perhaps *IRA1-missense* mutations
 433 have more diverse impacts on phenotype than do *IRA1-nonsense* mutations, which all likely
 434 result in destruction of the IRA1 protein. Our model also does not cluster the eight mutations in
 435 *IRA2* ($p = 0.086$) (Fig 3C; dark grey points). At the pathway level, our model does not cluster the
 436 three mutations to the TOR/Sch9 pathway away from the rest of the mutants, which are mainly
 437 in the Ras/PKA pathway ($p = 0.155$) (Fig 3C; purple points). Our model also does not cluster all
 438 diploids that possess additional mutations, including those with increased copy number of
 439 chromosome 11 or chromosome 12 and those with mutations in *IRA1* or *IRA2* ($p = 0.863$) (Fig
 440 3C; dark red points). Interestingly, our model does find a distinct cluster of six diploids that have
 441 higher than average diploid fitness in the EC ($p = 0.0001$) despite whole genome sequencing
 442 having revealed no mutations in their coding sequences (Fig 3C). This likely indicates that these
 443 diploids harbor difficult-to-sequence additional adaptive mutations that all have similar
 444 phenotypic consequences. In sum, these observations suggest that our genotype-phenotype-
 445 fitness model reveals new insights about which mutations affect the same functional units,
 446 specifically that these units do not always correspond to genes and pathways.

A

Measure fitness profiles of mutants across environments



Infer mutant phenotypes and environment weights from fitness profiles using SVD

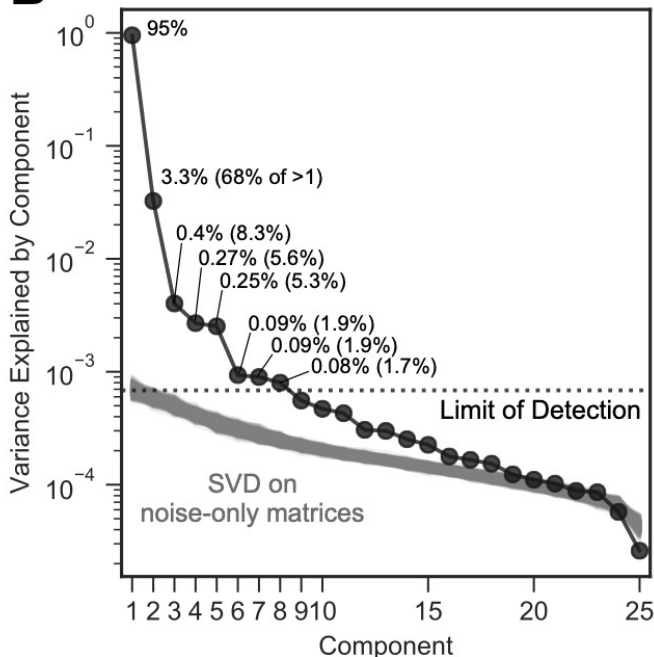
Mutant phenotypes (P)				Environment weights (E)				
P_{11}	P_{12}	\dots	P_{1k}	e_{11}	e_{12}	e_{13}	e_{14}	e_{15}
P_{21}	P_{22}	\dots	P_{2k}	e_{21}	e_{22}	e_{23}	e_{24}	e_{25}
P_{31}	P_{32}	\dots	P_{3k}	\vdots	\vdots	\vdots	\vdots	\vdots
P_{41}	P_{42}	\dots	P_{4k}	e_{k1}	e_{k2}	e_{k3}	e_{k4}	e_{k5}

$$f_{M_3E_2} = P_{31}e_{12} + P_{32}e_{22} + \dots + P_{3k}e_{k2}$$

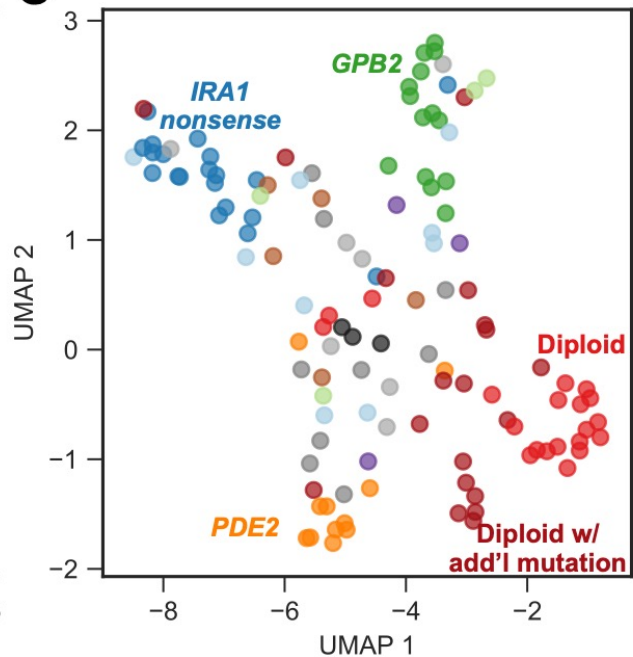
$$f_{M_4E_3} = P_{41}e_{13} + P_{42}e_{23} + \dots + P_{4k}e_{k3}$$

Linear model of phenotype to fitness

B



C



447 **Figure 3. Subtle environmental perturbations reveal an 8-component phenotypic model that**
448 **reflects known biological features. (A)** To infer fitness-relevant phenotypes, we measure the fitness of
449 mutants in a collection of environments and compare their fitness profiles. Mutants with similar fitness
450 profiles (mutants 1 and 2) are inferred to have similar effects on phenotypes. Mutants with dissimilar
451 fitness profiles (mutants 3 and 4) are inferred to have dissimilar phenotypic effects. We use SVD to
452 decompose these fitness profiles into a model consisting of two abstract spaces: one that represents the
453 fitness-relevant phenotypes affected by mutants (P) and another which represents the degree to which
454 each phenotype impacts fitness in each environment (E). Here, we represent the model with k fitness-
455 relevant phenotypes. The model's estimate for fitness for a particular mutant in a particular environment is
456 a linear combination of each mutant phenotype (mutant 1 is represented by the vector
457 $(p_{11}, p_{12}, p_{13}, \dots, p_{1k})$) scaled by the degree to which that phenotype affects fitness in the relevant
458 environment (environment 1 is represented by the vector $(e_{11}, e_{12}, e_{13}, \dots, e_{1k})$). We show two examples of
459 the equation used to estimate fitness for the mutants and environments highlighted in the left panel. Note
460 that, for presentation purposes, we show SVD as inferring two matrices. It in fact infers three, but is
461 consistent with our presentation if you fold the third matrix, which represents the singular values, into E
462 (see Methods). **(B)** Decomposing the fitness profiles of 292 adaptive mutants across 25 subtle
463 environmental perturbations reveals 8 fitness-relevant phenotypic components. The variance explained
464 by each component is indicated as a percentage of the total variance. The percentages in parentheses
465 indicate the relative amount of variation explained by each component when excluding the first
466 component. Each of these components explain more variation in fitness than do components that capture
467 variation across a simulated dataset in which fitness varies due to measurement noise. These simulations
468 were repeated 1000 times (grey lines) and used to define the limit of detection (dotted line). **(C)** An
469 abstract space containing 8 fitness-relevant phenotypic components reflects known biological features.
470 This plot shows the relationships of the mutants in a 7-dimensional phenotypic space that excludes the
471 first component, visualized using Uniform Manifold Approximation and Projection (UMAP). Mutants that
472 are close together have similar fitness profiles and are inferred to have similar effects on fitness-relevant
473 phenotypes. Mutants with mutations in the same gene tend to be closer together than random, in
474 particular *IRA1 nonsense* mutants in dark blue, *GPB2* mutants in dark green, *PDE2* mutants in dark
475 orange, and diploid mutants in red. Six diploid mutants that had higher than average diploid EC fitness
476 (and thus are likely to harbor additional mutation(s) so are categorized as “diploid with additional
477 mutation”) also form a cluster.

478 479 **Fitness variation across subtly different environments predicts fitness in substantially** 480 **different environments**

481
482 Now that we have identified the phenotypic components that contribute to fitness in
483 environments that represent subtle perturbations of the EC, we can test the ability of these
484 phenotypic components to predict fitness in more distant environments. Specifically, we can
485 measure how the contribution of each of these components to fitness changes in new
486 environments. We can also determine whether the phenotypic components that contribute very
487 little to explaining fitness variation near the EC might at times have large explanatory power in
488 distant environments (as depicted in Fig 1B).

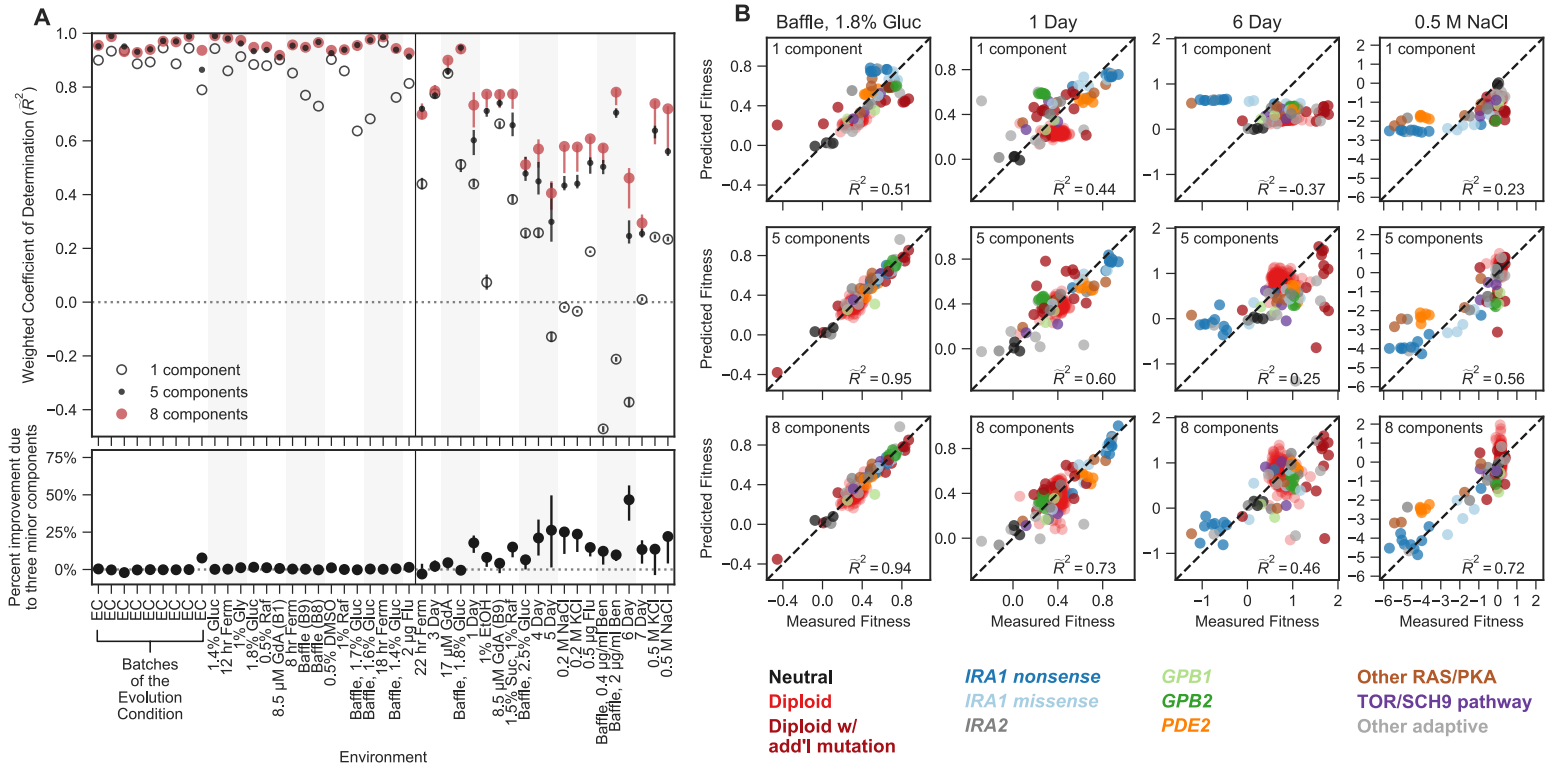
489
490 To test this we performed bi-cross-validation, using the eight component model constructed
491 from fitness variation of 60 training mutants across 25 subtly different environments to predict
492 the fitness of 232 test mutants in the environments that represent strong perturbations of the
493 EC. To evaluate the predictive power of the model, we compare our model's fitness predictions
494 in each environment to predictions made using the average fitness in that environment. Thus,
495 negative prediction power indicates cases where the model predicts fitness worse than
496 predictions using this average (Fig 4A).

497
498 The 8-dimensional phenotypic model, which was generated exclusively with the data from
499 subtle environmental perturbations, has substantial predictive power in distant environments
500 (Figure 4). Predictions explain 29% to 95% of the variation in fitness of the 232 test mutants
501 across strong environmental perturbations. For instance, in an environment where glucose
502 concentration was increased from 1.5% to 1.8% and the flask was changed to one that

503 increases the oxygenation of the media (the “Baffle, 1.8% Glucose” environment), we predict
504 95% of weighted variance with the full 8-component phenotypic model, in contrast to 51% with a
505 1-component model (Fig 4B). This ability to predict fitness is retained even when the first
506 component (effectively the fitness in EC) is a poor predictor of mutant fitness. For example, in
507 the environment where salt (0.5 M NaCl) was added to the media, the 1-component model
508 predicts fitness worse than predictions based on the average fitness for this environment,
509 resulting in negative variance explained (Fig 4A and 4B). This is due to the fact that mutant
510 fitness in this environment reflects extensive genotype-by-environment interactions, such that
511 the fitness of mutants in this environment is uncorrelated with EC fitness. However, our
512 predictions of mutant fitness in the 0.5 M NaCl environment improve when made using the 8-
513 component phenotypic model, which predicts 72% of weighted variance. Astoundingly, the 8-
514 component model captures strong tradeoffs between mutants with high fitness in the EC and
515 very low fitness in this high salt environment, specifically for *IRA1-nonsense* and, to a lesser
516 extent, *PDE2* mutants (Fig 4B). This was surprising because there appears to be very little
517 variation in fitness of these mutants across the subtle compared to the strong perturbations (Fig
518 2C).

519
520 This ability to predict fitness is also observed for mutations in genes and pathways that are not
521 represented in the 60 that comprise the training set (e.g. those with mutations in TOR/Sch9 and
522 HOG pathway genes). For example, the 8-component model explains 93% of variation in the
523 “Baffle, 1.8% Glucose” environment and 71% of variation in the 0.5M NaCl environment for
524 these mutations, compared to 76% and 31% variance explained for the 1-component model,
525 respectively. This indicates that our model is able to capture shared phenotypic effects that
526 extend beyond gene identity. Altogether, our ability to accurately predict the fitness of new
527 mutants in new environments suggests that the phenotypes our model identifies reflect causal
528 effects on fitness.

529
530 Most strikingly, phenotypic models that include the three smallest phenotypic components,
531 which together contribute only 1.1% to variance explained across the subtle environmental
532 perturbations (Fig 4A), often explain a substantial amount of variance in the distant
533 environments (Fig 4A; lower panel). For example, the three minor components contribute 17%
534 of the overall weighted variance explained in the 1-Day condition ($\bar{R}^2 = 0.6$ - 5-component
535 model, $\bar{R}^2 = 0.73$ - 8-component model; $(0.73-0.6)/0.73 = 0.17$) and 45% in the 6-Day
536 environment, ($\bar{R}^2 = 0.25$ - 5-component model, $\bar{R}^2 = 0.46$ - 8-component model) (Fig 4A and 4B).
537 In contrast, for other strong environments (e.g. Baffle - 1.8% Glucose, 8.5uM GdA (B9) and
538 Baffle - 2.5% Glucose), the three smallest components do not add much explanatory power (Fig
539 4A). These observations demonstrate that phenotypic components that make very small
540 contributions to fitness in the EC can contribute substantially to fitness in other environments.
541 Overall, these observations suggest an answer to questions about how adaptation is possible
542 when mutations have collateral effects on multiple phenotypes: not all of those phenotypes
543 contribute substantially to fitness in the EC (Fig 1B).



544 **Figure 4. Mutant fitness variation across subtly different environments predicts mutant fitness in**
 545 **novel and substantially different environments. (A) Top panel** vertical axis shows the accuracy of
 546 fitness predictions in each of 45 environments on the horizontal axis. The accuracy is calculated as the
 547 coefficient of determination, weighted such that each mutation type contributes equally. The left side of
 548 this plot represents predictions of mutant fitness in subtle environmental perturbations. These predictions
 549 are generated by holding out data from that environment when building the phenotypic model. The right
 550 side of the plot displays predictions of mutant fitness in strong environmental perturbations. These
 551 predictions are generated using a phenotypic model inferred from fitness variation across all 25 subtle
 552 different environments (denoted by each of the points or open circles) and for each of the 25 leave-one-
 553 out models (range of predictions is depicted with the error bars surrounding each point or open circle).
 554 Predictions from the 8-component model (red point) are typically better than the 1-component mode
 555 (open circle) and sometimes better than the 5-component model (black point). **Bottom panel** vertical axis
 556 shows the percent of the 8-component model's improvement due to the three minor components
 557 (calculated by the percent difference between the 5- and 8- component models). The left side shows the
 558 improvement of the prediction in subtle environmental perturbations when that subtle perturbation was
 559 held out. The right side shows the improvement of the prediction in strong environmental perturbations
 560 when using the full model (dots) or the 25 leave-one-out models (the error bars represent the range of
 561 improvement). **(B)** For each subplot, the horizontal axis shows the measured fitness value. The vertical
 562 axis shows the predicted fitness value when predictions are made using the 1-component (top row), 5-
 563 component (middle row), or 8-component (bottom row) models. Columns represent different
 564 environments. Points are colored by the mutation type. Note that \bar{R}^2 less than zero indicates that the
 565 prediction is worse than predictions using the mean fitness in that condition (see Methods).

566 **Not all mutants affect all phenotypes and not all environments make all phenotypes** 567 **important**

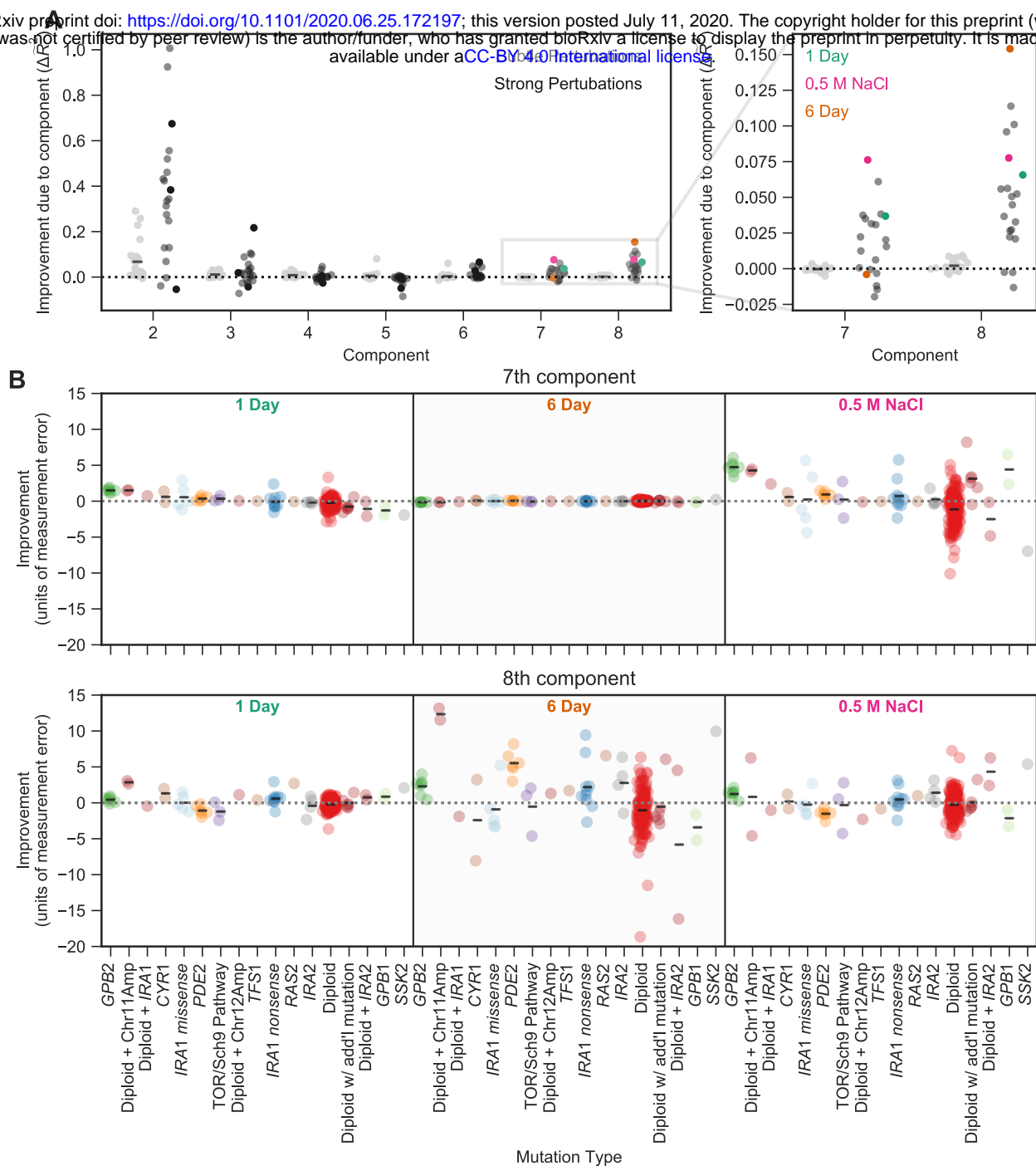
568
569 Next we explore the extent to which the contribution of a phenotypic component to fitness is
570 isolated to a specific environment and/or a specific type of mutation (Fig 5). We find that many
571 phenotypic components matter more to fitness in some environments than others. For instance,
572 component 2 adds on average 36% of the weighted variance in fitness across strong
573 perturbations, despite adding only 7% on average across the subtle environmental
574 perturbations. This contribution is, however, variable, with the second component adding over
575 90% of variance explained for the two environments with Benomyl and Baffled flasks (the
576 “Baffle, 0.4 $\mu\text{g/ml}$ Benomyl” and “Baffle, 2 $\mu\text{g/ml}$ Benomyl” environments) and only 0.3% for the
577 environment in which the transfer time was lengthened from two to three days (Fig 5A).

578
579 This environment-dependence is also true for the smallest two components. Specifically,
580 predictions of mutant fitness in the 0.5 M NaCl environment are improved from the inclusion of
581 component 7, adding 7.5% to weighted variance explained (Fig 5A). Predictions of mutant
582 fitness in the 6 Day transfer environment show improvement from the inclusion of the 8th
583 component, which adds over 15% to weighted variance explained (Fig 5A). However, the
584 predictions of fitness in the 6 Day environment are not improved from the inclusion of the 7th
585 component and the predictions in 0.5 M NaCl are not improved markedly by the inclusion of the
586 8th component (Fig 5A). This suggests that the phenotypic effects represented by these small
587 components contribute substantially in some environments and not others.

588
589 We further asked whether these effects are not only environment-specific but also mutant-
590 specific. To do so, we focused on environments for which the two smallest components
591 contribute substantially to fitness (e.g. 0.5 M NaCl). We looked at the extent to which each of
592 these components improves power to predict the fitness of each of the 232 held-out mutants.
593 We found these components improve the fitness predictions for some classes of mutants far
594 more than for others. For example, fitness predictions for mutations in *GPB2*, diploids with
595 chromosome 11 amplifications, and high-fitness diploids with no known mutations each
596 improved by over 4 standard deviations of measurement error in the 0.5 M NaCl environment
597 due to the inclusion of the 7th component (Fig 5B). This phenotypic component also has
598 importance in the 1-Day transfer environment, albeit to a lesser degree, resulting in
599 improvements of roughly 1 standard deviation for each of these mutation types. This suggests
600 that these mutants have some phenotypic effect that contributes only slightly to fitness in many
601 environments, including those that represent subtle perturbations of the EC, but that are
602 particularly important in the 0.5 M NaCl and 1 Day transfer environments. Similarly, we find that
603 the 8th component also improves predictive power for specific types of mutants in specific
604 environments. In this case, diploids with chromosome 11 amplifications and *PDE2* mutants have
605 particularly strong improvements in the 6-Day transfer environment (11 and 5 standard
606 deviations, respectively) and thus likely have a shared phenotypic effect that is captured by
607 component 8 (Fig 5B).

608
609 In sum, not all mutants affect all eight phenotypic components to the same degree and not
610 all phenotypic components contribute substantially to fitness in all environments. This
611 idiosyncrasy suggests that directional selection has the potential to generate rather than reduce
612 phenotypic diversity in cases where multiple adaptive mutants persist within a population or
613 across populations. Although directional selection “chooses” mutations that affect similar
614 phenotypes relevant to fitness in the EC, these mutations may have latent effects on a larger
615 number of phenotypes. When the environment changes, these latent phenotypic effects are
616 revealed, exposing the phenotypic diversity generated by the adaptive process.

617
618



619 **Figure 5. The contribution of a phenotypic component to fitness changes across environments**
 620 **and differs for different types of mutants. (A)** Some phenotypic components improve fitness
 621 predictions in some environments substantially more than they do in others. The vertical axis shows the
 622 improvement in the predictive power of our 8-component phenotypic model due to the inclusion of each
 623 component. For example, the improvement due to component 7 is calculated by the difference between
 624 the 7-component model and the 6-component model. The improvement of predictive power for each of
 625 the subtle environmental perturbations is shown as a gray point and for each of the strong perturbations
 626 in black. Magnification shows improvement upon including each of the two smallest components, with
 627 three strong perturbations highlighted. **(B)** Some phenotypic components improve fitness predictions for
 628 some mutants substantially more than they do for others. For example, the 7th component explains little
 629 variation in the 6-Day environment, but the 8th component explains a lot of variation in fitness in the 6-
 630 Day environment and is particularly helpful in predicting the fitness of Diploid + Chromosome 11
 631 Amplification mutations in this environment. Vertical axis shows the improvement in predictive power (in
 632 units of standard deviation of measurement error) for each type of mutant (denoted on the horizontal axis)
 633 in one of three environments (1 Day, 6 Day, and 0.5 M NaCl) when adding either the 7th (top panel) or
 634 the 8th (bottom panel) component. Mutants are ordered by the improvement due to the 7th component in
 635 the 1 Day environment. Since some types of mutants are more common, e.g. diploids, there are more
 636 data points in that category.

637 DISCUSSION

638
639 Here we succeeded in building a low-dimensional statistical model that captures the relationship
640 from genotype to phenotype to fitness for hundreds of adaptive mutants. Mapping the complete
641 phenotypic and fitness impacts of genetic change is a key goal of biology. Such a map is
642 important in order to make meaningful predictions from genetic data (e.g. personalized
643 medicine) and to investigate the structure of biological systems (e.g. their degree of modularity
644 and pleiotropy) (Collet et al., 2018; Eguchi et al., 2019; Exposito-Alonso et al., 2019; Zan and
645 Carlborg, 2020). Our model allows us to do both of these things. We made accurate predictions
646 about the fitness of unstudied mutants across multiple environments, and we gained novel
647 insights about the degree to which adaptive mutations are modular versus pleiotropic.
648 Specifically, we learned that adaptation is modular in the sense that hundreds of diverse
649 adaptive mutants collectively influence a small number of phenotypes that matter to fitness in
650 the evolution condition. We also learned that different mutants have distinct pleiotropic side
651 effects that matter to fitness in other conditions.

652
653 Building genotype-phenotype-fitness maps of adaptation has long been an elusive goal due to
654 both conceptual and technical difficulties. Indeed, the very first part of this task, namely the
655 identification of causal adaptive mutations, presents a substantial technical challenge (Barrett et
656 al., 2019, 2008; Exposito-Alonso et al., 2019). Fortunately, in some systems, such as in
657 microbial experimental evolution and studies of cancer and resistance in microbes and viruses,
658 genomic methodologies combined with availability of repeated evolutionary trials allow us to
659 detect specific genetic changes responsible for adaptation. In the context of microbial evolution
660 experiments, lineage tracing and genomics have opened up the possibility of not only detecting
661 hundreds of specific adaptive events but also measuring their fitness precisely and in bulk
662 (Good et al., 2017; Levy et al., 2015; Li et al., 2019, 2018; Nguyen Ba et al., 2019; Venkataram
663 et al., 2016). Thus in these cases we are coming close to solving the technical challenge of
664 building the *genotype to fitness* map of adaptation.

665
666 However, adding *phenotype* into this map remains a huge challenge even despite substantial
667 progress in mapping genotype to phenotype (Burga et al., 2019; Camp et al., 2019; Exposito-
668 Alonso et al., 2018; Geiler-Samerotte et al., 2016; Jakobson and Jarosz, 2019; Lee et al., 2019;
669 Paaby et al., 2015; Yengo et al., 2018; Ziv et al., 2017). In principle, we now have advanced
670 tools to measure a large number of phenotypic impacts of a genetic change, for instance
671 through high-throughput microscopy, proteomics, or RNAseq (Manzoni et al., 2018; Ritchie et
672 al., 2015; Zhang and Kuster, 2019). The conceptual problem is how to define phenotypes given
673 the interconnectedness of biological systems (Geiler-Samerotte et al., 2019; Paaby and
674 Rockman, 2013). If a mutation leads to complex changes in cell size and shape, should each
675 change be considered a distinct phenotype? Or if a single mutation changes the expression of
676 hundreds or thousands of genes, should we consider each change as a separate phenotype?
677 Intuitively, it seems that we should seek higher order, more meaningful descriptions. For
678 example, perhaps these expression changes are coordinated and reflect the up-regulation of a
679 stress-response pathway. Unfortunately, defining the functional units in which a gene product
680 participates remains difficult, especially because these units re-wire across genetic
681 backgrounds, environments, and species (Geiler-Samerotte et al., 2019; Pavličev et al., 2017;
682 Sun et al., 2020; Zan and Carlborg, 2020).

683
684 If mutations influence more than one phenotype, then the mapping from phenotype to fitness
685 also becomes challenging. To investigate this map, we would need to find an artificial way to
686 perturb one phenotype without perturbing others such that we could isolate and measure effects
687 on fitness. Mapping phenotype to fitness is further complicated by the environmental
688 dependence of these relationships (Fragata et al., 2019; Price et al., 2018). For example, a

689 mutation that affects a cell's ability to store carbohydrates for future use might matter far more in
690 an environment where glucose is re-supplied every 6 days instead of every 48 hours.

691
692 In our study, we turned the challenge of environment-dependence into the solution to the
693 seemingly intractable problem of interrogating the phenotype layer of the genotype-phenotype-
694 fitness map. We rely on the observation that the relative fitness of different mutations changes
695 across environments. We assume that differences in how mutant fitness varies across
696 environments must stem from differences in the phenotypes each mutation affects. Rather than
697 *a priori* defining the phenotypes that we think may matter, we use the similarities and
698 dissimilarities in the way fitness of multiple mutants vary across environments to define
699 phenotypes abstractly via their causal effects on fitness. This allows us to dispense with
700 measuring the phenotypes themselves and instead focus on measuring fitness with high
701 precision and throughput, since tools for doing so already exist (Venkataram et al., 2016). This
702 approach has the disadvantage of not identifying phenotypes in a traditional, more transparent
703 way. Still, it represents a major step forward in building genotype-phenotype-fitness maps
704 because it makes accurate predictions and provides novel insights about the phenotypic
705 structure of the adaptive response.

706
707 We successfully implemented this approach using a large collection of adaptive mutants
708 evolved in a glucose-limited condition. The first key result is that the map from adaptive mutant
709 to phenotype to fitness is modular, such that it is possible to create a genotype to phenotype to
710 fitness model that is low dimensional. Indeed, our model detects a small number (8) of fitness-
711 relevant phenotypes, the first two of which explain almost all of the variation in fitness (98.3%)
712 across 60 adaptive mutants in 25 environments representing subtle perturbations of the
713 glucose-limited evolution condition. This suggests that the hundreds of adaptive mutations we
714 study — including mutations in multiple genes in the Ras/PKA and TOR/Sch9 pathways,
715 genome duplication (diploidy), and various structural mutations — influence a small number of
716 phenotypes that matter to fitness in the evolution condition. This observation is consistent with
717 theoretical considerations suggesting that mutations that affect a large number of fitness-
718 relevant phenotypes are not likely to be adaptive (Orr, 2000; Wagner and Altenberg, 1996). It
719 also explains findings from other high-replicate laboratory evolution experiments and studies of
720 cancer that show hundreds of unique adaptive mutations tend to hit the same genes and
721 pathways repeatedly (Hanahan and Weinberg, 2011, 2000; Sanchez-Vega et al., 2018;
722 Tenaillon et al., 2012; Venkataram et al., 2016). Our work confirms the intuition that these
723 mutations all affect similar higher-order phenotypes (*e.g.* the level of activity of a signalling
724 pathway). This suggests that, despite the genetic diversity among adaptive mutants, adaptation
725 may be predictable and repeatable at the phenotypic level.

726
727 Note that although we detect only 8 fitness-relevant phenotypes, we expect the true number to
728 be much larger as the detectable number is limited by the precision of measurement (see
729 Methods and Fig S1). We expect this partly because we know that if we had worse precision in
730 this experiment we would have detected fewer than 8 phenotypic components (Fig 3). Still,
731 these additional undetected components cannot be very consequential in terms of their
732 contribution to fitness in the evolution condition, given how well the first 8 components capture
733 variation in environments that are similar to the evolution condition.

734
735 Surprisingly, the model built only using subtle environmental perturbations was also predictive of
736 fitness in environments that perturbed fitness strongly. Indeed in some of these environments,
737 such as the environment where 0.5 M NaCl was added to the media or the time of transfer was
738 extended from two to six days, many of the mutants are no longer adaptive and some of them
739 become strongly deleterious. Here the fitness of the mutants in the evolving condition is a very
740 poor predictor of fitness but the full 8-dimensional phenotypic model explains from 29% to 95%
741 of the variance. What was particularly interesting is that the explanatory power of different

742 dimensions was very different for the strong compared to subtle perturbations. For instance, the
743 second dimension which explained 7% of weighted variation on average in the subtle
744 perturbations, explained 36% on average in the environments that represent strong
745 perturbations. The pattern was particularly striking for the smallest dimensions which at times
746 explained 15% in the strong environmental perturbations while again explaining at most 1% in
747 the subtle environments.

748
749 This discovery emphasizes that, although the smaller phenotypic dimensions contribute very
750 little to fitness in the evolution condition, they can at times have a much larger contribution in
751 other environments (Fig 1B). This makes intuitive sense. For instance, we know that some of
752 the strongest adaptive mutations in our experiment, the nonsense mutations in *IRA1*, appear to
753 stop cells from shifting their metabolism towards carbohydrate storage when glucose levels
754 become low (Li et al., 2018). This gives these cells a head start once glucose again becomes
755 abundant and does not appear to come at a substantial cost, at least not until these cells are
756 exposed to stressful environments (e.g. high salt or long stationary phase) (Li et al., 2018). This
757 example, and more generally the observation that phenotypic effects that are unimportant in the
758 evolving condition can become much more important in other environments, supports the idea
759 that adaptation can happen through large effect mutations because many of the pleiotropic
760 phenotypic effects will be inconsequential in the local environment (Fig 1B). We can thus argue
761 that our low-dimensional model representing the genotype-phenotype-fitness map near the
762 evolution condition hides *latent* and *consequential* phenotypic complexity across the collection
763 of adaptive mutants. This complexity is hidden from natural selection in the evolution condition
764 but becomes important once the mutants leave the local environment and are assessed globally
765 for fitness effects. Thus, with respect to their effects on fitness-relevant phenotypes, adaptive
766 mutants may be locally modular, but globally pleiotropic.

767
768 The notion of latent phenotypic complexity is exciting as it generates a mechanism by which
769 directional selection generates rather than removes phenotypic diversity. Though directional
770 selection may promote multiple mutants that affect similar fitness-relevant phenotypes in the
771 evolution condition, each mutant could have disparate latent phenotypic effects that do not
772 contribute immediately to fitness. When the environment changes, these disparate phenotypic
773 effects may be revealed, imposing fitness costs of different magnitudes or allowing for diverse
774 solutions to a variety of possible new environments (Bono et al., 2017; Chavhan et al., 2020;
775 Jerison et al., 2020; Li et al., 2019). This latent phenotypic complexity also has the potential to
776 alter the future adaptive paths that a population takes even in a constant environment. Indeed,
777 these phenotypically diverse mutants are likely to affect the subsequent direction of adaptation
778 given that subsequent mutations can shift the context in which phenotypes are important in the
779 same way as do environmental perturbations (Blount et al., 2018, 2008; Dillon et al., 2016).
780 Latent phenotypic complexity among adaptive mutations is thus similar to cryptic genetic
781 variation in that it can influence a population's ability to adapt to new conditions (Paaby and
782 Rockman, 2013), but dissimilar in that it evolves under directional rather than stabilizing
783 selection. The end result is that directional selection can enhance diversity both within a
784 population in which multiple adaptive mutants are segregating and across populations that are
785 adapting to the same stressors.

786
787 The phenomenon of latent phenotypic complexity being driven by adaptation is dependent on
788 there being multiple mutational solutions to an environmental challenge, such that different
789 adaptive mutations might have different latent phenotypic effects. Latent phenotypic diversity
790 might be less apparent in cases where adaptation proceeds through mutations in a single gene
791 and certainly would not exist if adaptation relies on one unique mutation. Thus, in some ways
792 latent phenotypic diversity reflects redundancies in the mechanisms that allow cells to adapt to a
793 challenge. One such putative redundancy in the case investigated in this paper is that the
794 Ras/PKA pathway can be constitutively activated by loss of function mutations to a number of

795 negative regulators including *IRA1*, *PDE2*, and *GPB2*. Mutations in these genes might be
796 redundant in the sense that they influence the same fitness-relevant phenotype in the evolution
797 condition, which in this case is likely flux through the Ras/PKA pathway. This type of
798 redundancy is commonly observed in laboratory evolutions (Barghi et al., 2020), and is
799 particularly apparent in studies that analyze individuals with several adaptive mutations. Such
800 studies find that multiple mutations in the same functional unit occur less than expected by
801 chance presumably because those mutations would have redundant effects on fitness
802 (Tenaillon et al., 2012). Similarly, studies also find that second-step adaptive mutations tend to
803 be in different pathways or functional modules than the first adaptive step (Aggeli et al., 2020;
804 Fumasoni and Murray, 2020). The novel observation from our paper is that mutations with
805 redundant effects on fitness in the evolution condition are not necessarily identical because they
806 may influence different latent phenotypes. This observation adds to a long list of examples
807 demonstrating that redundancies, such as gene duplications and dominance, allow evolution the
808 flexibility to generate diversity.

809
810 One disadvantage of our approach is that the phenotypic components that we infer from our
811 fitness measurements are abstract. They represent causal effects on fitness, rather than
812 measurable features of cells. For this reason, perhaps we should not refer to them as
813 phenotypes but rather “fitnotypes” (a mash of the terms “fitness” and “phenotype”) that act much
814 like the causal traits in Fisher’s geometric model (Blanquart et al., 2014; Blanquart and
815 Bataillon, 2016; Fisher, 1930; Harmand et al., 2017; Lourenço et al., 2011; Martin and
816 Lenormand, 2006; Poon and Otto, 2000; Tenaillon, 2014; Tenaillon et al., 2007; Weinreich and
817 Knies, 2013) or a selectional pleiotropy model (Paaby and Rockman, 2013). Despite this
818 limitation, these fitnotypes have proven useful in allowing us to understand the consequences of
819 adaptive mutation. In addition to insights discussed above, we also learned that adaptive
820 mutants in the same gene do not always affect the same fitnotypes. For example, we found that
821 *IRA1-missense* mutations have varied and distinct effects from *IRA1-nonsense* mutations.
822 Further, we believe that identifying fitnotypes will ultimately prove useful in identifying the
823 phenotypic effects of mutation. The fitnotypes can serve as a scaffold onto which a large
824 number of phenotypic measurements can be mapped. Even though fitnotypes are independent
825 with respect to their contribution of fitness and contribute to fitness linearly, the mapping of
826 commonly measured features of cells (e.g. growth rate, the expression levels of growth
827 supporting proteins like ribosomes) onto fitnotypes may not be entirely straightforward.
828 Nonetheless, methods such as Sparse Canonical Correlation Analysis (Suo et al., 2017) hold
829 promise in such a mapping and might help us relate traditional phenotypes to fitnotypes.

830
831 An important question for future research is whether our observation of local modularity and
832 global pleiotropy are also apparent in other cases of adaptation. The method we described is
833 generic and can be applied to any system as long as the fitness of a substantial set of mutants
834 can be profiled across multiple environments or genetic backgrounds. This is becoming possible
835 to do in many systems (Jerison et al., 2020; Li et al., 2019; Martin et al., 2015; Pan et al., 2018;
836 Rogers et al., 2018). The notion that diverse genetic changes can have redundant effects in one
837 environment but distinct and consequential effects in other environments is important to our
838 understanding of adaptation in other settings, including in the context of antibiotic resistance
839 and cancer. For example, tumors representing the same type of cancer (e.g. lung
840 adenocarcinoma) tend to be genetically diverse even if considering only driver mutations (The
841 Cancer Genome Atlas Research Network, 2014). However, the driver mutations often fall into a
842 smaller number of key driver genes and even fewer pathways (Bailey et al., 2018; Hanahan and
843 Weinberg, 2011, 2000; Sanchez-Vega et al., 2018; Sondka et al., 2018). While this apparent
844 redundancy might suggest that the tumors are functionally similar, the notion of latent diversity
845 we propose here suggests that the specific mutational paths taken by different tumors might
846 matter once the environment changes, for example when the tumors are treated by a cancer
847 therapy. Substantial heterogeneity of tumor response to therapy is consistent with this notion.

848 Despite the accumulation of large amounts of genomic and phenomic data, integrating this
849 information to identify the phenotypic consequences of mutation that are ultimately responsible
850 for fitness remains incredibly challenging. Our approach allows us to create an abstract
851 representation of the causal effects of genetic mutation and their changing contribution to fitness
852 across environments. This top-down view of the genotype-phenotype-fitness map simplifies the
853 complex and multifaceted phenotypic consequences of mutation by focusing on those that
854 contribute to fitness. Integrating this new perspective with the influx of precise and high-
855 throughput data might allow us to answer age-old questions about the structure of biological
856 systems and adaptation.

857

858 **ACKNOWLEDGMENTS**

859 The authors thank Sandeep Venkataram for the BarcodeCounter2 script; Yuping Li, Monica
860 Sanchez, Tuya Yokoyama, Chris McFarland, and Dimitra Aggeli for technical assistance; Atish
861 Agarwala, Marc Salit, Sasha Levy, Gavin Sherlock, Ben Good, Ivana Cvijovic, David Gokhman,
862 Emily Ebel, Simon Levin, Molly Schumer, Jan Skotheim, Moises Exposito-Alonso, Mikhail
863 Tikhonov, Hunter Fraser, Michael Desai and all members of the Petrov Lab for helpful
864 comments and discussions. We are grateful to the twitter community that followed #1BigBatch
865 and provided us with very helpful feedback. We are grateful to Enrico Coen for very helpful
866 discussions and specifically for the suggestion of the term “fitnotype”. Some of the computing for
867 this project was performed on the Sherlock cluster. We would like to thank Stanford University
868 and the Stanford Research Computing Center for providing computational resources and
869 support that contributed to these research results. This work was supported by National
870 Institutes of Health grant R35GM118165 (to DAP) and National Institutes of Health grant
871 R35GM133674 (to KGS).

872 REFERENCES

- 873 Aggeli, D., Li, Y., Sherlock, G., 2020. Changes in the distribution of fitness effects and adaptive
874 mutational spectra following a single first step towards adaptation (preprint). *Evolutionary*
875 *Biology*. <https://doi.org/10.1101/2020.06.12.148833>
- 876 Altenberg, L., 2005. Modularity in Evolution: Some Low-Level Questions, in: *Modularity:*
877 *Understanding the Development and Evolution of Complex Natural Systems*. MIT Press,
878 p. 32.
- 879 Baeza-Centurion, P., Miñana, B., Schmiedel, J.M., Valcárcel, J., Lehner, B., 2019.
880 *Combinatorial Genetics Reveals a Scaling Law for the Effects of Mutations on Splicing*.
881 *Cell* 176, 549-563.e23. <https://doi.org/10.1016/j.cell.2018.12.010>
- 882 Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A.,
883 Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al., 2018. Comprehensive
884 Characterization of Cancer Driver Genes and Mutations. *Cell* 173, 371-385.e18.
885 <https://doi.org/10.1016/j.cell.2018.02.060>
- 886 Barghi, N., Hermisson, J., Schlötterer, C., 2020. Polygenic adaptation: a unifying framework to
887 understand positive selection. *Nat. Rev. Genet.* [https://doi.org/10.1038/s41576-020-](https://doi.org/10.1038/s41576-020-0250-z)
888 [0250-z](https://doi.org/10.1038/s41576-020-0250-z)
- 889 Barghi, N., Tobler, R., Nolte, V., Jakšić, A.M., Mallard, F., Otte, K.A., Dolezal, M., Taus, T.,
890 Kofler, R., Schlötterer, C., 2019. Genetic redundancy fuels polygenic adaptation in
891 *Drosophila*. *PLOS Biol.* 17, e3000128. <https://doi.org/10.1371/journal.pbio.3000128>
- 892 Barrett, R.D.H., Laurent, S., Mallarino, R., Pfeifer, S.P., Xu, C.C.Y., Foll, M., Wakamatsu, K.,
893 Duke-Cohan, J.S., Jensen, J.D., Hoekstra, H.E., 2019. Linking a mutation to survival in
894 wild mice. *Science* 363, 499–504. <https://doi.org/10.1126/science.aav3824>
- 895 Barrett, R.D.H., Rogers, S.M., Schluter, D., 2008. Natural Selection on a Major Armor Gene in
896 Threespine Stickleback. *Science* 322, 255–257. <https://doi.org/10.1126/science.1159978>
- 897 Blanquart, F., Achaz, G., Bataillon, T., Tenaillon, O., 2014. Properties of selected mutations and
898 genotypic landscapes under Fisher's geometric model: GENOTYPIC LANDSCAPES
899 UNDER FISHER'S MODEL. *Evolution* 68, 3537–3554.
900 <https://doi.org/10.1111/evo.12545>
- 901 Blanquart, F., Bataillon, T., 2016. Epistasis and the Structure of Fitness Landscapes: Are
902 Experimental Fitness Landscapes Compatible with Fisher's Geometric Model? *Genetics*
903 203, 847–862. <https://doi.org/10.1534/genetics.115.182691>
- 904 Blount, Z.D., Borland, C.Z., Lenski, R.E., 2008. Historical contingency and the evolution of a key
905 innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci.* 105,
906 7899–7906. <https://doi.org/10.1073/pnas.0803151105>
- 907 Blount, Z.D., Lenski, R.E., Losos, J.B., 2018. Contingency and determinism in evolution:
908 Replaying life's tape. *Science* 362, eaam5979. <https://doi.org/10.1126/science.aam5979>
- 909 Bono, L.M., Smith, L.B., Pfennig, D.W., Burch, C.L., 2017. The emergence of performance
910 trade-offs during local adaptation: insights from experimental evolution. *Mol. Ecol.* 26,
911 1720–1733. <https://doi.org/10.1111/mec.13979>
- 912 Boyle, E.A., Li, Y.I., Pritchard, J.K., 2017. An Expanded View of Complex Traits: From
913 Polygenic to Omnigenic. *Cell* 169, 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- 914 Burga, A., Ben-David, E., Lemus Vergara, T., Boocock, J., Kruglyak, L., 2019. Fast genetic
915 mapping of complex traits in *C. elegans* using millions of individuals in bulk. *Nat.*
916 *Commun.* 10, 2680. <https://doi.org/10.1038/s41467-019-10636-9>
- 917 Camp, J.G., Platt, R., Treutlein, B., 2019. Mapping human cell phenotypes to genotypes with
918 single-cell genomics. *Science* 365, 1401–1405. <https://doi.org/10.1126/science.aax6648>
- 919 Chavhan, Y., Malusare, S., Dey, S., 2020. Larger bacterial populations evolve heavier fitness
920 trade-offs and undergo greater ecological specialization. *Heredity* 124, 726–736.
921 <https://doi.org/10.1038/s41437-020-0308-x>
- 922 Chesmore, K., Bartlett, J., Williams, S.M., 2018. The ubiquity of pleiotropy in human disease.
923 *Hum. Genet.* 137, 39–44. <https://doi.org/10.1007/s00439-017-1854-z>

- 924 Collet, J.M., McGuigan, K., Allen, S.L., Chenoweth, S.F., Blows, M.W., 2018. Mutational
925 Pleiotropy and the Strength of Stabilizing Selection Within and Between Functional
926 Modules of Gene Expression. *Genetics* 208, 1601–1616.
927 <https://doi.org/10.1534/genetics.118.300776>
- 928 Coombes, D., Moir, J.W.B., Poole, A.M., Cooper, T.F., Dobson, R.C.J., 2019. The fitness
929 challenge of studying molecular adaptation. *Biochem. Soc. Trans.* 47, 1533–1542.
930 <https://doi.org/10.1042/BST20180626>
- 931 Cowperthwaite, M.C., Bull, J.J., Meyers, L.A., 2005. Distributions of Beneficial Fitness Effects in
932 RNA. *Genetics* 170, 1449–1457. <https://doi.org/10.1534/genetics.104.039248>
- 933 Crozat, E., Winkworth, C., Gaffe, J., Hallin, P.F., Riley, M.A., Lenski, R.E., Schneider, D., 2010.
934 Parallel Genetic and Phenotypic Evolution of DNA Superhelicity in Experimental
935 Populations of *Escherichia coli*. *Mol. Biol. Evol.* 27, 2113–2128.
936 <https://doi.org/10.1093/molbev/msq099>
- 937 Dillon, M.M., Rouillard, N.P., Van Dam, B., Gallet, R., Cooper, V.S., 2016. Diverse phenotypic
938 and genetic responses to short-term selection in evolving *Escherichia coli* populations:
939 DIVERSE RESPONSES TO SELECTION OF *E. coli* POPULATIONS. *Evolution* 70,
940 586–599. <https://doi.org/10.1111/evo.12868>
- 941 Diss, G., Lehner, B., 2018. The genetic landscape of a physical interaction. *eLife* 7, e32472.
942 <https://doi.org/10.7554/eLife.32472>
- 943 Domingo, J., Baeza-Centurion, P., Lehner, B., 2019. The Causes and Consequences of Genetic
944 Interactions (Epistasis). *Annu. Rev. Genomics Hum. Genet.* 20, 433–460.
945 <https://doi.org/10.1146/annurev-genom-083118-014857>
- 946 Eckart, C., Young, G., 1936. The approximation of one matrix by another of lower rank.
947 *Psychometrika* 1, 211–218. <https://doi.org/10.1007/BF02288367>
- 948 Eguchi, Y., Bilolikar, G., Geiler-Samerotte, K., 2019. Why and how to study genetic changes
949 with context-dependent effects. *Curr. Opin. Genet. Dev.* 58–59, 95–102.
950 <https://doi.org/10.1016/j.gde.2019.08.003>
- 951 Exposito-Alonso, M., 500 Genomes Field Experiment Team, Burbano, H.A., Bossdorf, O.,
952 Nielsen, R., Weigel, D., 2019. Natural selection on the *Arabidopsis thaliana* genome in
953 present and future climates. *Nature* 573, 126–129. <https://doi.org/10.1038/s41586-019-1520-9>
- 954
- 955 Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H.A., Weigel, D., 2018.
956 Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis*
957 *thaliana*. *Nat. Ecol. Evol.* 2, 352–358. <https://doi.org/10.1038/s41559-017-0423-0>
- 958 Fisher, R.A., 1930. The genetical theory of natural selection. Clarendon Press, Oxford.
959 <https://doi.org/10.5962/bhl.title.27468>
- 960 Fragata, I., Blanckaert, A., Dias Louro, M.A., Liberles, D.A., Bank, C., 2019. Evolution in the
961 light of fitness landscape theory. *Trends Ecol. Evol.* 34, 69–82.
962 <https://doi.org/10.1016/j.tree.2018.10.009>
- 963 Fumasoni, M., Murray, A.W., 2020. The evolutionary plasticity of chromosome metabolism
964 allows adaptation to constitutive DNA replication stress. *eLife* 9, e51963.
965 <https://doi.org/10.7554/eLife.51963>
- 966 Geiler-Samerotte, K.A., Li, S., Lazaris, C., Taylor, A., Ziv, N., Ramjeawan, C., Paaby, A.B.,
967 Siegal, M.L., 2019. Extent and context dependence of pleiotropy revealed by high-
968 throughput single-cell phenotyping (preprint). *Evolutionary Biology*.
969 <https://doi.org/10.1101/700716>
- 970 Geiler-Samerotte, K.A., Zhu, Y.O., Goulet, B.E., Hall, D.W., Siegal, M.L., 2016. Selection
971 Transforms the Landscape of Genetic Variation Interacting with Hsp90. *PLOS Biol.* 14,
972 e2000465. <https://doi.org/10.1371/journal.pbio.2000465>
- 973 Good, B.H., McDonald, M.J., Barrick, J.E., Lenski, R.E., Desai, M.M., 2017. The dynamics of
974 molecular evolution over 60,000 generations. *Nature* 551, 45–50.
975 <https://doi.org/10.1038/nature24287>
- 976 Hanahan, D., Weinberg, R.A., 2011. Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–

- 977 674. <https://doi.org/10.1016/j.cell.2011.02.013>
- 978 Hanahan, D., Weinberg, R.A., 2000. The Hallmarks of Cancer. *Cell* 100, 57–70.
- 979 [https://doi.org/10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9)
- 980 Harmand, N., Gallet, R., Jabbour-Zahab, R., Martin, G., Lenormand, T., 2017. Fisher's
- 981 geometrical model and the mutational patterns of antibiotic resistance across dose
- 982 gradients: FISHER'S GEOMETRICAL MODEL AND ANTIBIOTIC RESISTANCE.
- 983 *Evolution* 71, 23–37. <https://doi.org/10.1111/evo.13111>
- 984 Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W., 1999. From molecular to modular cell
- 985 biology. *Nature* 402, C47–C52. <https://doi.org/10.1038/35011540>
- 986 Huang, C.-J., Lu, M.-Y., Chang, Y.-W., Li, W.-H., 2018. Experimental Evolution of Yeast for
- 987 High-Temperature Tolerance. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msy077>
- 988 Illumina, 2017. Effects of Index Misassignment on Multiplexing and Downstream Analysis
- 989 [WWW Document]. URL [https://www.illumina.com/content/dam/illumina-](https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf?linkId=36607862)
- 990 [marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-](https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf?linkId=36607862)
- 991 [004.pdf?linkId=36607862](https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf?linkId=36607862). (accessed 6.25.20).
- 992 Jakobson, C.M., Jarosz, D.F., 2019. Molecular Origins of Complex Heritability in Natural
- 993 Genotype-to-Phenotype Relationships. *Cell Syst.* 8, 363-379.e3.
- 994 <https://doi.org/10.1016/j.cels.2019.04.002>
- 995 Jerison, E.R., Nguyen Ba, A.N., Desai, M.M., Kryazhimskiy, S., 2020. Chance and necessity in
- 996 the pleiotropic consequences of adaptation for budding yeast. *Nat. Ecol. Evol.* 4, 601–
- 997 611. <https://doi.org/10.1038/s41559-020-1128-3>
- 998 Josse, J., Sardy, S., 2014. Adaptive Shrinkage of singular values. *ArXiv13106602 Stat.*
- 999 Kacser, H., Burns, J.A., 1981. The Molecular Basis of Dominance. *Genetics* 97, 639–666.
- 1000 Karageorgi, M., Groen, S.C., Sumbul, F., Pelaez, J.N., Verster, K.I., Aguilar, J.M., Hastings,
- 1001 A.P., Bernstein, S.L., Matsunaga, T., Astourian, M., et al., 2019. Genome editing
- 1002 retraces the evolution of toxin resistance in the monarch butterfly. *Nature* 574, 409–412.
- 1003 <https://doi.org/10.1038/s41586-019-1610-8>
- 1004 Kemble, H., Eisenhauer, C., Couce, A., Chapron, A., Magnan, M., Gautier, G., Le Nagard, H.,
- 1005 Nghe, P., Tenaillon, O., 2020. Flux, toxicity, and expression costs generate complex
- 1006 genetic interactions in a metabolic pathway. *Sci. Adv.* 6, eabb2236.
- 1007 <https://doi.org/10.1126/sciadv.abb2236>
- 1008 Lang, G.I., Rice, D.P., Hickman, M.J., Sodergren, E., Weinstock, G.M., Botstein, D., Desai,
- 1009 M.M., 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast
- 1010 populations. *Nature* 500, 571–574. <https://doi.org/10.1038/nature12344>
- 1011 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
- 1012 9, 357–359. <https://doi.org/10.1038/nmeth.1923>
- 1013 Lee, J.T., Coradini, A.L.V., Shen, A., Ehrenreich, I.M., 2019. Layers of Cryptic Genetic Variation
- 1014 Underlie a Yeast Complex Trait. *Genetics* 211, 1469–1482.
- 1015 <https://doi.org/10.1534/genetics.119.301907>
- 1016 Levy, S.F., Blundell, J.R., Venkataram, S., Petrov, D.A., Fisher, D.S., Sherlock, G., 2015.
- 1017 Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519,
- 1018 181–186. <https://doi.org/10.1038/nature14279>
- 1019 Li, C., Zhang, J., 2018. Multi-environment fitness landscapes of a tRNA gene. *Nat. Ecol. Evol.* 2,
- 1020 1025–1032. <https://doi.org/10.1038/s41559-018-0549-8>
- 1021 Li, Y., Petrov, D.A., Sherlock, G., 2019. Single nucleotide mapping of trait space reveals Pareto
- 1022 fronts that constrain adaptation. *Nat. Ecol. Evol.* 3, 1539–1551.
- 1023 <https://doi.org/10.1038/s41559-019-0993-0>
- 1024 Li, Y., Venkataram, S., Agarwala, A., Dunn, B., Petrov, D.A., Sherlock, G., Fisher, D.S., 2018.
- 1025 Hidden Complexity of Yeast Adaptation under Simple Evolutionary Conditions. *Curr.*
- 1026 *Biol.* 28, 515-525.e6. <https://doi.org/10.1016/j.cub.2018.01.009>
- 1027 Lind, P.A., Farr, A.D., Rainey, P.B., 2015. Experimental evolution reveals hidden diversity in
- 1028 evolutionary pathways. *eLife* 4, e07074. <https://doi.org/10.7554/eLife.07074>
- 1029 Lourenço, J., Galtier, N., Glémin, S., 2011. COMPLEXITY, PLEIOTROPY, AND THE FITNESS

- 1030 EFFECT OF MUTATIONS. *Evolution* 65, 1559–1571.
- 1031 Manzoni, C., Kia, D.A., Vandrovcova, J., Hardy, J., Wood, N.W., Lewis, P.A., Ferrari, R., 2018.
- 1032 Genome, transcriptome and proteome: the rise of omics data and their integration in
- 1033 biomedical sciences. *Brief. Bioinform.* 19, 286–302. <https://doi.org/10.1093/bib/bbw114>
- 1034 Marsaglia, G., 1972. Choosing a Point from the Surface of a Sphere. *Ann. Math. Stat.* 43, 645–
- 1035 646. <https://doi.org/10.1214/aoms/1177692644>
- 1036 Martin, G., Lenormand, T., 2006. A GENERAL MULTIVARIATE EXTENSION OF FISHER'S
- 1037 GEOMETRICAL MODEL AND THE DISTRIBUTION OF MUTATION FITNESS
- 1038 EFFECTS ACROSS SPECIES. *Evolution* 60, 893–907.
- 1039 Martin, H., Shales, M., Fernandez-Piñar, P., Wei, P., Molina, M., Fiedler, D., Shokat, K.M.,
- 1040 Beltrao, P., Lim, W., Krogan, N.J., 2015. Differential genetic interactions of yeast stress
- 1041 response MAPK pathways. *Mol. Syst. Biol.* 11, 800.
- 1042 <https://doi.org/10.15252/msb.20145606>
- 1043 McInnes, L., Healy, J., Melville, J., 2018. UMAP: Uniform Manifold Approximation and Projection
- 1044 for Dimension Reduction. *ArXiv180203426 Cs Stat.*
- 1045 Mehlhoff, J.D., Stearns, F.W., Rohm, D., Wang, B., Tsou, E.-Y., Dutta, N., Hsiao, M.-H.,
- 1046 Gonzalez, C.E., Rubin, A.F., Ostermeier, M., 2020. Collateral fitness effects of
- 1047 mutations. *Proc. Natl. Acad. Sci.* 117, 11597–11607.
- 1048 <https://doi.org/10.1073/pnas.1918680117>
- 1049 Melo, D., Porto, A., Cheverud, J.M., Marroig, G., 2016. Modularity: Genes, Development, and
- 1050 Evolution. *Annu. Rev. Ecol. Evol. Syst.* 47, 463–486. <https://doi.org/10.1146/annurev-ecolsys-121415-032409>
- 1051 Nguyen Ba, A.N., Cvijović, I., Rojas Echenique, J.I., Lawrence, K.R., Rego-Costa, A., Liu, X.,
- 1052 Levy, S.F., Desai, M.M., 2019. High-resolution lineage tracking reveals travelling wave of
- 1053 adaptation in laboratory yeast. *Nature* 575, 494–499. [https://doi.org/10.1038/s41586-](https://doi.org/10.1038/s41586-019-1749-3)
- 1054 [019-1749-3](https://doi.org/10.1038/s41586-019-1749-3)
- 1055
- 1056 Orr, H.A., 2000. Adaptation and The Cost of Complexity. *Evolution* 54, 13–20.
- 1057 <https://doi.org/10.1111/j.0014-3820.2000.tb00002.x>
- 1058 Otwinowski, J., McCandlish, D.M., Plotkin, J.B., 2018. Inferring the shape of global epistasis.
- 1059 *Proc. Natl. Acad. Sci.* 115, E7550–E7558. <https://doi.org/10.1073/pnas.1804015115>
- 1060 Owen, A.B., Perry, P.O., 2009. Bi-cross-validation of the SVD and the nonnegative matrix
- 1061 factorization. *Ann. Appl. Stat.* 3, 564–594. <https://doi.org/10.1214/08-AOAS227>
- 1062 Paaby, A.B., Rockman, M.V., 2013. The many faces of pleiotropy. *Trends Genet.* 29, 66–73.
- 1063 <https://doi.org/10.1016/j.tig.2012.10.010>
- 1064 Paaby, A.B., White, A.G., Riccardi, D.D., Gunsalus, K.C., Piano, F., Rockman, M.V., 2015. Wild
- 1065 worm embryogenesis harbors ubiquitous polygenic modifier variation. *eLife* 4, e09178.
- 1066 <https://doi.org/10.7554/eLife.09178>
- 1067 Pan, J., Meyers, R.M., Michel, B.C., Mashtalir, N., Sizemore, A.E., Wells, J.N., Cassel, S.H.,
- 1068 Vazquez, F., Weir, B.A., Hahn, W.C., et al., 2018. Interrogation of Mammalian Protein
- 1069 Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens.
- 1070 *Cell Syst.* 6, 555-568.e7. <https://doi.org/10.1016/j.cels.2018.04.011>
- 1071 Pavličev, M., Wagner, G.P., Chavan, A.R., Owens, K., Maziarz, J., Dunn-Fletcher, C., Kallapur,
- 1072 S.G., Muglia, L., Jones, H., 2017. Single-cell transcriptomics of the human placenta:
- 1073 inferring the cell communication network of the maternal-fetal interface. *Genome Res.*
- 1074 27, 349–361. <https://doi.org/10.1101/gr.207597.116>
- 1075 Poon, A., Otto, S.P., 2000. COMPENSATING FOR OUR LOAD OF MUTATIONS: FREEZING
- 1076 THE MELTDOWN OF SMALL POPULATIONS. *Evolution* 54, 1467–1479.
- 1077 <https://doi.org/10.1111/j.0014-3820.2000.tb00693.x>
- 1078 Pressman, A.D., Liu, Z., Janzen, E., Blanco, C., Müller, U.F., Joyce, G.F., Pascal, R., Chen,
- 1079 I.A., 2019. Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated
- 1080 Evolutionary Network for Self-Aminoacylating RNA. *J. Am. Chem. Soc.* 141, 6213–6223.
- 1081 <https://doi.org/10.1021/jacs.8b13298>
- 1082 Price, N., Moyers, B.T., Lopez, L., Lasky, J.R., Monroe, J.G., Mullen, J.L., Oakley, C.G., Lin, J.,

- 1083 Ågren, J., Schrider, D.R., et al., 2018. Combining population genomics and fitness QTLs
1084 to identify the genetics of local adaptation in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci.
1085 115, 5028–5033. <https://doi.org/10.1073/pnas.1719998115>
- 1086 Ritchie, M.D., Holzinger, E.R., Li, R., Pendergrass, S.A., Kim, D., 2015. Methods of integrating
1087 data to uncover genotype–phenotype interactions. Nat. Rev. Genet. 16, 85–97.
1088 <https://doi.org/10.1038/nrg3868>
- 1089 Rockman, M.V., 2012. THE QTN PROGRAM AND THE ALLELES THAT MATTER FOR
1090 EVOLUTION: ALL THAT’S GOLD DOES NOT GLITTER. Evolution 66, 1–17.
1091 <https://doi.org/10.1111/j.1558-5646.2011.01486.x>
- 1092 Rogers, Z.N., McFarland, C.D., Winters, I.P., Seoane, J.A., Brady, J.J., Yoon, S., Curtis, C.,
1093 Petrov, D.A., Winslow, M.M., 2018. Mapping the in vivo fitness landscape of lung
1094 adenocarcinoma tumor suppression in mice. Nat. Genet. 50, 483–486.
1095 <https://doi.org/10.1038/s41588-018-0083-2>
- 1096 Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu,
1097 D.L., Kantheti, H.S., Saghafeina, S., et al., 2018. Oncogenic Signaling Pathways in The
1098 Cancer Genome Atlas. Cell 173, 321-337.e10. <https://doi.org/10.1016/j.cell.2018.03.035>
- 1099 Sarkisyan, K.S., Bolotin, D.A., Meer, M.V., Usmanova, D.R., Mishin, A.S., Sharonov, G.V.,
1100 Ivankov, D.N., Bozhanova, N.G., Baranov, M.S., Soylemez, O., et al., 2016. Local fitness
1101 landscape of the green fluorescent protein. Nature 533, 397–401.
1102 <https://doi.org/10.1038/nature17995>
- 1103 Sella, G., Barton, N.H., 2019. Thinking About the Evolution of Complex Traits in the Era of
1104 Genome-Wide Association Studies. Annu. Rev. Genomics Hum. Genet. 20, 461–493.
1105 <https://doi.org/10.1146/annurev-genom-083115-022316>
- 1106 Sinha, R., Stanley, G., Gulati, G.S., Ezran, C., Travaglini, K.J., Wei, E., Chan, C.K.F., Nabhan,
1107 A.N., Su, T., Morganti, R.M., et al., 2017. Index switching causes “spreading-of-signal”
1108 among multiplexed samples in Illumina HiSeq 4000 DNA sequencing (preprint).
1109 Molecular Biology. <https://doi.org/10.1101/125724>
- 1110 Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J.G., Zgaga, L., Manolio, T.,
1111 Rudan, I., McKeigue, P., Wilson, J.F., Campbell, H., 2011. Abundant Pleiotropy in
1112 Human Complex Diseases and Traits. Am. J. Hum. Genet. 89, 607–618.
1113 <https://doi.org/10.1016/j.ajhg.2011.10.004>
- 1114 Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., Forbes, S.A., 2018. The COSMIC
1115 Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat.
1116 Rev. Cancer 18, 696–705. <https://doi.org/10.1038/s41568-018-0060-1>
- 1117 Starr, T.N., Flynn, J.M., Mishra, P., Bolon, D.N.A., Thornton, J.W., 2018. Pervasive contingency
1118 and entrenchment in a billion years of Hsp90 evolution. Proc. Natl. Acad. Sci. 115,
1119 4453–4458. <https://doi.org/10.1073/pnas.1718133115>
- 1120 Sun, S., Baryshnikova, A., Brandt, N., Gresham, D., 2020. Genetic interaction profiles of
1121 regulatory kinases differ between environmental conditions and cellular states. Mol.
1122 Syst. Biol. 16. <https://doi.org/10.15252/msb.20199167>
- 1123 Suo, X., Minden, V., Nelson, B., Tibshirani, R., Saunders, M., 2017. Sparse canonical
1124 correlation analysis. ArXiv170510865 Stat.
- 1125 Tenailon, O., 2014. The Utility of Fisher’s Geometric Model in Evolutionary Genetics. Annu.
1126 Rev. Ecol. Evol. Syst. 45, 179–201. <https://doi.org/10.1146/annurev-ecolsys-120213-091846>
- 1127
1128 Tenailon, O., Rodriguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D.,
1129 Gaut, B.S., 2012. The Molecular Diversity of Adaptive Convergence. Science 335, 457–
1130 461. <https://doi.org/10.1126/science.1212986>
- 1131 Tenailon, O., Silander, O.K., Uzan, J.-P., Chao, L., 2007. Quantifying Organismal Complexity
1132 using a Population Genetic Approach. PLoS ONE 2, e217.
1133 <https://doi.org/10.1371/journal.pone.0000217>
- 1134 The Cancer Genome Atlas Research Network, 2014. Comprehensive molecular profiling of lung
1135 adenocarcinoma. Nature 511, 543–550. <https://doi.org/10.1038/nature13385>

- 1136 Venkataram, S., Dunn, B., Li, Y., Agarwala, A., Chang, J., Ebel, E.R., Geiler-Samerotte, K.,
1137 Hérissant, L., Blundell, J.R., Levy, S.F., et al., 2016. Development of a Comprehensive
1138 Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell* 166, 1585-
1139 1596.e22. <https://doi.org/10.1016/j.cell.2016.08.002>
- 1140 Venkataram, S., Monasky, R., Sikaroodi, S.H., Kryazhimskiy, S., Kaçar, B., 2019. Evolutionary
1141 Stalling and a Limit on the Power of Natural Selection to Improve a Cellular Module
1142 (preprint). *Evolutionary Biology*. <https://doi.org/10.1101/850644>
- 1143 Verduyn, C., Postma, E., Scheffers, W.A., Van Dijken, J.P., 1992. Effect of benzoic acid on
1144 metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration
1145 and alcoholic fermentation. *Yeast* 8, 501–517. <https://doi.org/10.1002/yea.320080703>
- 1146 Visscher, P.M., Yang, J., 2016. A plethora of pleiotropy across complex traits. *Nat. Genet.* 48,
1147 707–708. <https://doi.org/10.1038/ng.3604>
- 1148 Wagner, G.P., Altenberg, L., 1996. Perspective: Complex Adaptations and the Evolution of
1149 Evolvability. *Evolution* 50, 967–976. <https://doi.org/10.2307/2410639>
- 1150 Wagner, G.P., Pavlicev, M., Cheverud, J.M., 2007. The road to modularity. *Nat. Rev. Genet.* 8,
1151 921–931. <https://doi.org/10.1038/nrg2267>
- 1152 Wagner, G.P., Zhang, J., 2011. The pleiotropic structure of the genotype–phenotype map: the
1153 evolvability of complex organisms. *Nat. Rev. Genet.* 12, 204–213.
1154 <https://doi.org/10.1038/nrg2949>
- 1155 Weinreich, D.M., 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to
1156 Fitter Proteins. *Science* 312, 111–114. <https://doi.org/10.1126/science.1123539>
- 1157 Weinreich, D.M., Knies, J.L., 2013. FISHER’S GEOMETRIC MODEL OF ADAPTATION MEETS
1158 THE FUNCTIONAL SYNTHESIS: DATA ON PAIRWISE EPISTASIS FOR FITNESS
1159 YIELDS INSIGHTS INTO THE SHAPE AND SIZE OF PHENOTYPE SPACE: THE FGM
1160 AND THE FUNCTIONAL SYNTHESIS. *Evolution* n/a-n/a.
1161 <https://doi.org/10.1111/evo.12156>
- 1162 Welch, J.J., Waxman, D., 2003. MODULARITY AND THE COST OF COMPLEXITY. *Evolution*
1163 57, 1723–1734. <https://doi.org/10.1111/j.0014-3820.2003.tb00581.x>
- 1164 Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M.,
1165 Hirschhorn, J., Yang, J., Visscher, P.M., et al., 2018. Meta-analysis of genome-wide
1166 association studies for height and body mass index in ~700000 individuals of European
1167 ancestry. *Hum. Mol. Genet.* 27, 3641–3649. <https://doi.org/10.1093/hmg/ddy271>
- 1168 Zan, Y., Carlborg, Ö., 2020. Dynamic genetic architecture of yeast response to environmental
1169 perturbation shed light on origin of cryptic genetic variation. *PLOS Genet.* 16, e1008801.
1170 <https://doi.org/10.1371/journal.pgen.1008801>
- 1171 Zhang, B., Kuster, B., 2019. Proteomics Is Not an Island: Multi-omics Integration Is the Key to
1172 Understanding Biological Systems. *Mol. Cell. Proteomics* 18, S1–S4.
1173 <https://doi.org/10.1074/mcp.E119.001693>
- 1174 Ziv, N., Shuster, B.M., Siegal, M.L., Gresham, D., 2017. Resolving the Complex Genetic Basis
1175 of Phenotypic Variation and Variability of Cellular Growth. *Genetics* 206, 1645–1657.
1176 <https://doi.org/10.1534/genetics.116.195180>
- 1177
1178
1179
1180
1181
1182
1183
1184

1185 **METHODS**

1186 **LEAD CONTACT AND MATERIALS AVAILABILITY**

1187 Further information and requests for resources and reagents should be directed to and will be
1188 fulfilled by the Lead Contact, Dmitri Petrov (dpetrov@stanford.edu).

1189 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

1190 The yeast strains used in this study can be grown and maintained using standard methods (e.g.
1191 YPD media in test tubes, glycerol stocks for long term storage at -80°C), but should be
1192 propagated in the appropriate selection environment (a glucose-limited minimal media - M3
1193 medium for the evolution condition) for comparable fitness and phenotypic measurements. All of
1194 the strains we study are of genetic background MAT α , *ura3 Δ 0*, *ybr209w::Gal-Cre-KanMX-*
1195 *1/2URA3-loxP-Barcode-1/2URA3-HygMX-lox66/71*.

1196 Experiments were performed with barcoded mutants isolated from a previous evolution
1197 experiment (Levy et al., 2015). To measure their fitness, these mutants were competed against
1198 a constructed reference strain with a restriction site in the barcode region (Venkataram et al.,
1199 2016).

1200 The majority of the fitness measurement experiments were conducted with a collection of 500
1201 adaptive barcoded mutants where each strain starts at equal frequency (Li et al., 2018;
1202 Venkataram et al., 2016). We focus on a subset of 292 strains for which we obtained fitness
1203 measurements in all 45 environments and for which mutations conferring fitness advantages
1204 have been previously identified, either by whole genome sequencing or using a drug to test
1205 ploidy (Li et al., 2018; Venkataram et al., 2016) (Table S1). Note that because we utilize some
1206 data from previous experiments (Li et al., 2018; Venkataram et al., 2016), some of the
1207 experiments contained additional barcoded mutants not analyzed here, namely a pool
1208 consisting of a total of 4800 strains, including the 292 focused on in this study. These
1209 differences in the number of strains included in the experiment are partially accounted for in our
1210 inference of mean fitness, and any remaining effects can be thought of as another parameter
1211 that varies across the environments (e.g. in addition to glucose or salt concentration).

1212 In a few experiments, we spiked in re-barcoded mutants and additional neutral lineages as
1213 internal controls. Since re-barcoded mutants are identical, except for the barcode, these teach
1214 us about the precision with which we can measure a mutant's fitness. Specifically, we spiked in
1215 ten re-barcoded *IRA1-nonsense* mutants (each with a frameshift insertion AT to ATT mutation at
1216 bp 4090) and ten *IRA1-missense* mutants (each with a G to T mutation at bp 3776). Neutral
1217 lineages teach us about the behavior of the unmutated reference strain, which we must infer
1218 because its barcode is eliminated from the experiment before sequencing. The spiked in
1219 neutrals include ten barcoded lineages from the original evolution experiment (Levy et al., 2015)
1220 for which whole genome sequencing did not reveal any mutations (Venkataram et al., 2016) and
1221 previous fitness measurements did not reveal any deviation from the reference (Li et al., 2018;
1222 Venkataram et al., 2016).

1223 **METHOD DETAILS**

1224 **Conducting the barcoded fitness measurements**

1225
1226 Fitness measurement experiments were performed as described previously (Li et al., 2018;
1227 Venkataram et al., 2016), where growth competitions were set up between a pool of barcoded
1228 mutants and a reference strain. The change in the frequency of each barcode over time reflects
1229 the fitness of the adaptive mutant possessing that barcode, relative to the reference strain.

1230
1231 We conducted fitness measurements under a variety of conditions (Table S2) that represent
1232 perturbations of the condition in which these adaptive mutants evolved. Briefly, we separately
1233 grew up an overnight culture of the barcode pool and the ancestral reference strain in 100mL
1234 M3 (minimal, glucose-limited) medium (Verduyn et al., 1992). We then mixed these saturated
1235 cultures at a 1:9 ratio such that 90% of cells represent the reference strain. This ratio allows for
1236 mutants to compete against the ancestor rather than competing against each other. We then
1237 inoculated 400 μ L of this mixed culture ($\sim 5 \times 10^7$ cells) into 100mL of fresh media in 500mL
1238 DeLong flasks. The type of media used, and sometimes the shape of the flask, varied
1239 depending on condition (Table S2). This culture was then grown at 30°C in an incubator shaking
1240 at 223 RPM for 48 hours. After 48 hours of growth, 400 μ L of saturated culture was transferred
1241 into fresh media of the same type, in a new flask of the same type. This serial dilution was
1242 usually continued 4 times, yielding 5 time-points over which to measure the rate at which a
1243 barcode's frequency changed, though some experiments include one more or one less
1244 depending on the experimenter and on whether technical problems (e.g. PCR failure) caused
1245 loss of time-points.

1246
1247 After each transfer of 400 μ L, the left-over 9600 μ L was frozen so that we could later sequence
1248 the barcodes present at every time-point. To prepare this culture for freezing, it was transferred
1249 to 50mL conicals, spun down at 3000 rpm for 5 minutes, resuspended in 5mL of sorbitol
1250 freezing solution (0.9M sorbitol, 0.1M Tris-HCL pH 7.5, 0.1M EDTA pH 8.0), aliquoted into three
1251 1.5mL tubes, and stored at -80°C.

1252
1253 For experiments where additional neutral lineages and re-barcoded lineages were included, the
1254 initial inoculation mix consisted of 90% ancestral reference strain, 9.4% barcode mutant pool,
1255 0.2% additional neutral spike-in pool, 0.2% re-barcoded IRA1 nonsense pool, and 0.2% re-
1256 barcoded IRA1 missense pool.

1257 **Growth conditions**

1258
1259 In this study, we present fitness measurement data from a collection of 45 conditions that each
1260 represent perturbations of the growth condition in which these adaptive mutants evolved. We
1261 refer to this original evolution condition as the "EC". In the EC, cells are grown in flasks with a
1262 flat bottom and transferred to new flasks every 48 hours (see *Conducting the barcoded fitness*
1263 *measurements*). Cells are grown in M3 media (Verduyn et al., 1992). This media is glucose-
1264 limited, meaning the cells run out of glucose before any other nutrient. In the EC, the starting
1265 glucose concentration is 1.5%.

1266
1267 The 45 perturbations of the EC are summarized in Table S2 and include changes to the growth
1268 media, the flask shape, and the transfer times. For example, in the "1 Day" condition, we
1269 change the transfer time from 48 to 24 hours. In the "1.8% glucose, baffled flask condition" we
1270 change the starting glucose concentration from 1.5% to 1.8% and change the flask type from
1271 one with a flat bottom to one with baffles. Several of these conditions include experiments from
1272 previous studies (Li et al., 2018; Venkataram et al., 2016).

1273

1274 For each of these 45 conditions but three, we include between two and four replicates that were
1275 performed simultaneously (Table S2) such that overall we performed a total of 109 fitness
1276 measurements on our collection of adaptive mutants. Our replicate structure is nested in that
1277 some of our 45 conditions represent replicate experiments that we performed at different times.
1278 Variation across experiments performed at different times is often referred to as “batch effects”
1279 and likely reflects environmental variability that we were unable to control (e.g. slight fluctuations
1280 in incubation temperature due to limits on the precision of the instrument). In particular, we re-
1281 measured the fitness of the adaptive mutants in the EC on 9 different occasions, each time
1282 including 3 or more replicates. We refer to these 9 experiments as ‘EC batches’ in the main text.
1283 However, every set of experiments that was performed at the same time constitutes a separate
1284 “batch”. There were slight differences across batches in the way we prepared barcodes for
1285 sequencing, which we detail in the relevant Methods sections. This variation across batches can
1286 be thought of as another parameter that varies across the 45 conditions (in addition to glucose
1287 or salt concentration). We report which experiments were performed in the same batch in Table
1288 S2.

1289
1290 Some conditions, including some Fluconazole conditions and Geldanamycin conditions, have
1291 unexpected orderings in the strength of perturbation (i.e. the smaller drug concentration shows
1292 a larger difference in fitness or similar concentrations seem to have different effects).
1293 Regardless of whether these observations reflect technical problems (e.g. degradation or poor
1294 solubility of the drug), we include these conditions because we use the effect of the realized
1295 perturbation on fitness to build low-dimensional phenotypic models. In other words, the identity
1296 of the perturbation does not matter in this study.

1297 **DNA Extraction of each sample**

1299 After a growth competition is complete, we extracted DNA from frozen samples following either
1300 a protocol described previously (for batches 1 – 6 and 10) (Venkataram et al., 2016) or a
1301 modified protocol that improves the ease and yield of extraction. Our modified protocol is as
1302 follows. For each sample, a single tube of the three that were frozen for each sample (see
1303 *Conducting the barcoded fitness measurements*) was removed from the freezer and thawed at
1304 room temperature. We extracted DNA from that sample using the following modification of the
1305 Lucigen MasterPure yeast DNA purification kit (#MPY80200). We transferred the thawed cells
1306 into a 15mL conical and centrifuge for 3 min at 4000 RPM. After discarding the supernatant, the
1307 pellet was then resuspended with 1.8 mL of the MasterPure lysis buffer, and 0.5 mm glass
1308 beads were added to help with disruption of the yeast cell wall. The mix of pellet, lysis buffer,
1309 and beads was then vortexed for 10 seconds and incubated for 45 minutes at 65°C, with
1310 periodic vortexing. The solution was then put on ice for 5 min and then 900 μ L of MPC Protein
1311 Reagent was mixed with the solution. We then separated protein and cell debris by
1312 centrifugation at 4000 RPM, transferring 1900 μ L of supernatant to a 2 mL centrifuge tube. We
1313 further separated remaining protein and cell debris by centrifuging at 13200 RPM for 5 min. The
1314 supernatant was then divided into two 2mL centrifuge tubes, with 925 μ L of the supernatant into
1315 each. Next, we added 1000 μ L of isopropanol to each tube, mixed by inversion, centrifuged at
1316 13200 RPM for 5 min, and discarded the supernatant. The pellet, containing the DNA was then
1317 resuspended in 250 μ L of Elution Buffer and 10 μ L of 5 ng/ μ L RNAase A was added. This was
1318 either left at room temperature overnight or incubated at 60°C for 15 min. Next the two tubes per
1319 sample were combined into a single tube and 1500 μ L of ethanol was added. This was then
1320 mixed by inversion, and strands of precipitating DNA appeared. This was centrifuged at 13200
1321 RPM for 2 min, and the supernatant was discarded. We again precipitated the DNA by
1322 resuspending with 750 μ L of ethanol, and collected the DNA by centrifuging 13200 RPM for 2
1323 min. The supernatant was discarded, and the tubes were left to air dry. Finally, we resuspended
1324 the pellet in Elution Buffer to a final concentration of 50 ng/ μ L for later use in PCR reactions
1325 (approximately 3600 ng of DNA were used for the PCR reactions).

1326

1327 PCR Amplification of the Barcode Locus

1328 After extracting DNA, we PCR-amplified the barcode locus for each sample. Batches 1 – 6 and
1329 10 were conducted with the protocols described in (Li et al., 2018; Venkataram et al., 2016). We
1330 made some slight modifications to this protocol, including using a new set of primers to allow for
1331 nested-unique-dual index labeling, for batches 7, 8, and 9. Our modified protocol is as follows.

1332
1333 We used a two-step PCR protocol to amplify the barcodes from the DNA. The first PCR cycle
1334 uses primers with “inline indices” to label samples (see *Mitigating the effects of index hopping*
1335 section for details). These inline indices are highlighted in bold below. Attaching unique indices
1336 to samples pertaining to different conditions or timepoints allows us to multiplex these samples
1337 on the same sequencing lane. Each primer also contains a Unique Molecular Identifier (UMI) –
1338 denoted by the sequence of “N” nucleotides in the primer – which is used to determine if
1339 identical barcode sequences each represent yeast cells that were present at the time the
1340 sample was frozen, or a PCR amplification of the a barcode from a single cell (see Levy et al.,
1341 2015; Li et al., 2018; Venkataram et al., 2016). Primers were HPLC purified to ensure they are
1342 the correct length.

1343 1344 Forward primers

Primer Name	Sequence
F201	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN CGATGTT TAATATGGACTAAAGGAGGCTTTT
F202	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN ACAGTGT TAATATGGACTAAAGGAGGCTTTT
F203	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN TGACCAT TAATATGGACTAAAGGAGGCTTTT
F204	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN GCCAATT TAATATGGACTAAAGGAGGCTTTT
F205	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN ATCACGT TAATATGGACTAAAGGAGGCTTTT
F206	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN CAGATCT TAATATGGACTAAAGGAGGCTTTT
F207	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN GGCTACT TAATATGGACTAAAGGAGGCTTTT
F208	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN TAGCTTT TAATATGGACTAAAGGAGGCTTTT
F209	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN TTAGGCT TAATATGGACTAAAGGAGGCTTTT
F210	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN ACTTGAT TAATATGGACTAAAGGAGGCTTTT
F211	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN GATCAGT TAATATGGACTAAAGGAGGCTTTT
F212	TCGTCGGCAGCGTC AGATGTGTATAAGAGACAG NNNNNNNN CTTGTAT TAATATGGACTAAAGGAGGCTTTT

1345 1346 Reverse primers

Primer Name	Sequence
R301	GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG NNNNNNNN TATATACGC TCGAATTCAAGCTTAGATCTGATA
R302	GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG NNNNNNNN CGCTCTATC TCGAATTCAAGCTTAGATCTGATA
R303	GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG NNNNNNNN GAGACGTCT TCGAATTCAAGCTTAGATCTGATA
R304	GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG NNNNNNNN ATACTGCGT TCGAATTCAAGCTTAGATCTGATA
R305	GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG NNNNNNNN ACTAGCAGA TCGAATTCAAGCTTAGATCTGATA
R306	GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG NNNNNNNN TGAGCTAGC TCGAATTCAAGCTTAGATCTGATA
R307	GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG NNNNNNNN CTGCTACTC TCGAATTCAAGCTTAGATCTGATA
R308	GTCTCGTGGGCTCGG AGATGTGTATAAGAGACAG NNNNNNNN GCGTACGCA TCGAATTCAAGCTTAGATCTGATA

1347
1348
1349 For the first step of PCR, we performed 8 reactions per sample to offset the effects of PCR
1350 jackpotting within each reaction. For each set of 8 reactions, we used the master mix:
1351 - 200 μ L OneTaq Hot Start 2X Master Mix with Standard Buffer (NEB M0484L)

- 1352 - 8 μ L 10uM Forward primer
- 1353 - 8 μ L 10uM Reverse primer
- 1354 - 72 μ L sample genomic DNA (diluted to 50ng/ μ L or all of sample if between 25-50ng/ μ L)
- 1355 - 16 μ L 50mM MgCl₂
- 1356 - 96 μ L Nuclease Free Water (Fisher Scientific #AM9937)

1357

1358 We then aliquoted 50 μ L of the master mix into each of 8 PCR tubes, and ran on the
1359 thermocycler with the following cycle:

- 1360 1. 94°C for 10 min
- 1361 2. 94°C for 3 min
- 1362 3. 55°C for 1 min
- 1363 4. 68°C for 1 min
- 1364 5. Repeat steps 2-4 2x (for a total of 3 cycles)
- 1365 6. 68°C for 1 min
- 1366 7. Hold at 4°C

1367

1368 We then added 100 μ L of binding buffer from the ThermoScientific GeneJET Gel Extraction Kit
1369 (#K0692) to each PCR reaction, and performed a standard PCR purification protocol in one
1370 column per sample. In the final step, we eluted into 80 μ L of elution buffer.

1371

1372 For the second step of PCR, we use standard Nextera XT Index v2 primers (Illumina #FC-131-
1373 2004) to further label samples representing different conditions and timepoints with unique
1374 identifiers that allow for multiplexing on the same sequencing lane. We uniquely dual-indexed
1375 each sample using our nested scheme (see *Mitigating the effects of index hopping* section for
1376 details). We performed 3 reactions of the second step PCR per sample, using the master mix:

- 1377 - 1.5 μ L Q5 Polymerase (NEB #M0491L)
- 1378 - 30 μ L Q5 Buffer (NEB #M0491L)
- 1379 - 3 μ L 10mM dNTP (Fisher Scientific #PR-U1515)
- 1380 - 6.25 μ L i7 Nextera XT Primer (“N” primer)
- 1381 - 6.25 μ L i5 Nextera XT Primer (“S” primer)
- 1382 - 78 μ L purified step 1 PCR product
- 1383 - 25 μ L Nuclease Free Water (Fisher Scientific #AM9937)

1384

1385 This master mix was then divided into 3 PCR tubes per reaction, and run with the following
1386 protocol on a thermocycler:

1387

- 1388 1. 98°C for 30 sec
- 1389 2. 98°C for 10 sec
- 1390 3. 62°C for 20 sec
- 1391 4. 72°C for 30 sec
- 1392 5. Repeat steps 2-4 at least 21 times and at most 27 times (for a total of 22 to 28 cycles)
- 1393 6. 72°C for 3 min
- 1394 7. Hold at 4°C

1395

1396 We then added 100 μ L of binding buffer from the ThermoScientific GeneJET Gel Extraction Kit
1397 and purified the PCR product, eluting into 43 μ L. We found that increasing the number of cycles
1398 in the second step PCR beyond 21 did not seem to improve the amount of DNA recovered after
1399 gel extraction. For some samples, we experimented with a touch down procedure for the
1400 second step PCR where we started with a hotter annealing temperature and slowly decreased it
1401 over the course of 27 cycles. This also did not seem to increase the yield of DNA recovered
1402 from the PCR.

1403

1404 **Removal of the Reference Strain via Digestion and Gel Purification**

1405 To avoid the vast majority of our sequencing reads mapping only to the reference strain (and
1406 thus not being informative to relative fitness of the mutants), we use restriction digest to cut the
1407 ApaLI restriction site in the middle of the reference strain's barcode region. We mixed 43 μ L of
1408 the second step PCR product with 2 μ L of ApaLI (NEB #R0507L) and 5 μ L of 10X Cutsmart and
1409 incubated at 37°C for at least 2 hours (up to overnight). After digestion, we conducted size
1410 selection by running the digested sample on a gel, removing all product less than 300bp, and
1411 isolating the DNA using a standard ThermoScientific GeneJET Gel Extraction protocol. Our
1412 expected product is 350bp. We did not remove longer sequences via gel extraction because of
1413 the possibility that some barcode sequences may selectively form complexes with themselves
1414 or other barcodes.

1415
1416 Note that for some samples, we also digested the reference strain before PCR, in addition to
1417 after PCR, to decrease the amount of reference strain barcode. For these samples, we mixed
1418 80ul of genomic DNA (at concentration 50ng/ μ L) with 10 μ L of 10X Cutsmart and 2 μ L of ApaLI
1419 and incubated 37°C for at least 2 hours (up to overnight). This product was then used as the
1420 template for PCR step 1 (with appropriate water volume adjustments to ensure 50 μ L reactions).

1421 1422 **Sample pooling and Amplicon Sequencing**

1423 We used the Qubit High Sensitivity (ThermoFisher #Q32854) method to quantify the
1424 concentration of the final product for each sample, then pooled samples with different dual
1425 indices in equal frequency for sequencing. Our samples were then sent to either Novogene
1426 (<https://en.novogene.com/>) or Admera Health (<https://www.admerahealth.com/>) for quality
1427 control (qPCR and either Bioanalyzer or TapeStation) and sequencing. We used 2x150 paired-
1428 end sequencing along with index sequencing reads on Illumina HiSeq machines using patterned
1429 flow cells (either HiSeq 4000 or HiSeq X). We also used Illumina Nextseq machines with
1430 unpatterned flow cells. We found that the former was more subject to index hopping errors,
1431 please see *Mitigating the effects of index hopping* for a discussion of how our dual indexing
1432 reduces effects of index hopping. All amplicon samples were sequenced with at least 20%
1433 genomic DNA spiked in (either whole genomes from an unrelated project or phi-X) to ensure
1434 adequate diversity on the flow cell.

1435 1436 **Mitigating the effects of index hopping**

1437 To reduce the effects of index hopping observed on Illumina patterned flow cell technology
1438 (including HiSeq 4000, HiSeq X, and Novaseq machines) (Illumina, 2017; Sinha et al., 2017),
1439 we devise a nested unique-dual-indexing approach. This approach uses a combination of inline
1440 indices attached during the first step of PCR, as well as Nextera indices attached during the
1441 second step of PCR. The latter indices are not part of the sequencing read (they are read in a
1442 separate Index Read). This process uniquely labels both ends of all DNA strands such that DNA
1443 strands from multiple samples can be multiplexed on the same flow cell. Had we only labeled
1444 one end of each DNA strand, index hopping could have caused us to incorrectly identify some
1445 reads as coming from the wrong sample.

1446
1447 One approach to label samples with unique-dual-indices is to use 96 forward primers, each of
1448 which is paired to one of 96 reverse primers, instead our nested approach allows us to uniquely
1449 dual-index samples with only 40 total primers (12 forward inline, 8 reverse inline, 12 Nextera i7,
1450 8 Nextera i5). Specifically, we can use combinations of the Nextera and inline primers. One way
1451 to think of this is that there are 96 possible ways to combine the forward inline and Nextera i5
1452 primers that are on the same side of the read, effectively creating 96 unique labels for that end
1453 of the read.

1454
1455 To reduce the effect of index hopping contamination on our results, we included only samples
1456 that were sequenced on non-patterned flow cell technology (HiSeq 2000 and 2500 for samples

1457 in batches 1-6, 10, NextSeq for samples in batch 9) or were sequenced on patterned flow cell
1458 technology (patterned flow cell HiSeq) with nested unique-dual indexing.

1459

1460 **Processing of Amplicon Sequencing Data**

1461 We processed the amplicon sequencing data by first using the index tags to de-multiplex reads
1462 representing different conditions and timepoints. Then, using Bowtie2 (Langmead and Salzberg,
1463 2012), we mapped reads to a known list of barcodes generated by Venkataram et al. (2016),
1464 removed PCR duplicates using the UMIs from the first-step primers, and counted the number of
1465 reads for each barcode in each sample. The source code for this step can be found at
1466 <https://github.com/sandeepvenkataram/BarcodeCounter2>. We processed all raw data for this
1467 study using this pipeline, including re-processing the raw sequencing files for data from previous
1468 studies (Li et al., 2018; Venkataram et al., 2016) so that all data was processed together using
1469 the most recent version of the code.

1470

1471 Several samples included technical replicates where the sample was split at various times in the
1472 process, including before DNA extraction, before PCR, and prior to sequencing. Read counts
1473 across these technical replicates were merged in order to calculate the best estimate of barcode
1474 frequencies. Counts were merged after appropriately accounting for PCR duplicates as
1475 identified from Unique Molecular Identifiers.

1476

1477

1478 **QUANTIFICATION AND STATISTICAL ANALYSIS**

1479

1480 **Fitness Estimate Inference**

1481 The amplicon sequencing data shows the relative frequency of each barcode in each time-point
1482 of every one of our 109 fitness measurement experiments. To estimate the fitness of each
1483 barcoded mutant in each experiment, we calculate how barcode frequencies change over time.
1484 We do this using previously described methods (Venkataram et al., 2016).

1485

1486 Briefly, we first calculated the log-frequency change of each barcoded adaptive mutant for each
1487 subsequent pair of time-points. This log-frequency change must be corrected by the mean
1488 fitness of the population, such that it represents the relative fitness of each mutant relative to the
1489 reference strain, which makes up the bulk of the population. Since we destroyed barcodes
1490 pertaining to the reference strain by digesting them, we infer how the mean fitness of the
1491 population changes at each time-point using barcoded lineages that are known to be neutral
1492 (see *Identification of neutral lineages*). Once we calculated the change in the relative fitness of
1493 each barcoded mutant across each pair of consecutive time-points, we took a weighted average
1494 across all pairs as our final estimate of each adaptive mutant's relative fitness for a given
1495 experiment. We weighted each pairwise fitness estimate using an uncertainty measure
1496 generated from a noise model (see *Noise model* section below).

1497

1498 This results in 109 fitness measurements per each barcoded mutant, with some of the 45
1499 conditions having more representation than others due to having more replicates. In cases
1500 where we have replicates, we averaged the fitness values across the replicates, weighted by
1501 the measurement uncertainty, resulting in our final 45 fitness estimates per each adaptive
1502 mutant lineage.

1503

1504 We included only timepoints with at least 1,000 reads for which at least 400 mutants were
1505 detected to have at least 1 read. Furthermore, we required that fitness measurement
1506 experiments must include at least three timepoints to be included in our analysis.

1507

1508 **Identification of Neutral Lineages**

1509 Previous work using this fitness measurement method focused on a larger collection of 4800
1510 barcoded yeast lineages, where the vast majority of these lineages were neutral (Li et al., 2018;
1511 Venkataram et al., 2016). In order to increase the number of reads per adaptive lineage, we
1512 used a smaller pool of 500 lineages for most experiments. However, this prevents us from
1513 identifying neutral lineages as was done in previous studies, by rejecting outlier lineages with
1514 higher than typical fitness values. Instead, we used a set of 35 high-confidence neutral lineages
1515 to infer mean fitness (see *Experimental model and subject details*). These lineages showed no
1516 fitness differences from the neutral expectation in previous studies and were shown to possess
1517 no mutations detectable via whole genome sequencing. These high-confidence neutral lineages
1518 were present in all experiments, and were spiked into experiments from batch 9 to increase their
1519 frequency. We used these neutrals to perform the fitness inference in two steps. First, we
1520 inferred fitness using this collection of high-confidence neutrals to make a first pass at inferring
1521 the fitness values. Next, we included lineages with similar behavior to the high-confidence
1522 neutrals to improve our estimate of mean fitness.

1523 1524 **Noise model**

1525 To quantify the uncertainty for each fitness measurement, we used the noise model as outlined
1526 in Venkataram et al., 2016.

1527
1528 Briefly, this noise model accounts for the uncertainty coming from several sources of noise. The
1529 first type of noise scales with the number of reads for a given lineage. This noise stems from
1530 stochasticity in population dynamics (coming from the inherent stochasticity in growth and noise
1531 associated with dilution), from counting noise associated with a finite coverage, and technical
1532 noise from DNA extraction and PCR. We fit this noise by quantifying the variation in the
1533 frequency of neutral lineages (see *Identification of neutral lineages*). There is additional variation
1534 in fitness observed for high-frequency lineages between replicate experiments (here we refer to
1535 variation across replicates that were performed simultaneously, not variation across batches).
1536 We also accounted for this uncertainty following previous studies. Specifically, we fit an
1537 additional frequency-independent source of noise using between-replicate variation.

1538 1539 **Checks on noise model**

1540 Because our ability to count the phenotypes that matter to fitness hinges upon measurement
1541 error, we further assessed the accuracy of our noise model. We did so by using barcoded
1542 lineages that should have the same fitness because they are genetically identical. Since our
1543 fitness estimates are imperfect (*i.e.* they contain some noise), we estimated each of these
1544 lineages as having slightly different fitness. We then asked if the variation in fitness across
1545 these lineages is explained by our noise model, or if there is more variation than our noise
1546 model can account for. We did this explicitly by calculating, for each lineage, how far our fitness
1547 estimate is from the best guess for the true underlying fitness value (the group's mean) in units
1548 of the estimate's measurement precision. We then calculated the percent of lineages that are a
1549 given distance from the group's average to understand the accuracy of the model. For instance,
1550 if the noise model perfectly captures the uncertainty of each measurement, then 10% of the
1551 diploid lineages should have a difference from the weighted diploid mean in the 10th percentile,
1552 20% in the 20th percentile, etc. Because 188 of our 292 barcode mutants are diploids without
1553 additional mutations, diploids are an ideal group to use to assess the accuracy of the noise
1554 model. This procedure shows that, for the vast majority of replicates, the noise model is
1555 conservative. That is, diploid lineages tend to have less variation in fitness than expected by the
1556 noise model (Fig S1).

1557 1558 **Classifying mutants by mutation type**

1559 Some types of mutants are present more than others. For example, 188 of our 292 mutants are
1560 diploids and 30 mutants are in the IRA1 gene. If not properly accounted for, this imbalance can
1561 lead to some unfairness in predictions for our model. For example, if we use mostly diploid

1562 lineages to train our model, we will be very good at predicting the fitness of diploids but poor at
1563 predicting other types of mutants. This means that we must classify our mutants by mutation
1564 type in order to properly balance them. We classified mutants following previous work
1565 (Venkataram et al., 2016) that classified mutants as either diploids, or if haploid, by the gene
1566 possessing the putative causal mutation. Because previous work finds differences in fitness
1567 between missense and nonsense/frameshift/indel mutations in IRA1, here we classified these
1568 mutants into “missense” and “nonsense” classes, where mutants with frameshift and indel
1569 mutations were classified as “nonsense”. We also classified diploid mutants with additional
1570 mutations in nutrient-response genes or chromosomal amplifications as separate groups.
1571 Additionally, we created a separate class for “high-fitness diploid” mutants that possess no
1572 additional detected mutations (other than being diploid) but have very high fitness in the EC. To
1573 be classified as a high-fitness diploid, a diploid mutant must have an average fitness across all 9
1574 EC batches that is greater than 2 standard deviations above the average of all diploids. In the
1575 main text, we label these mutants as “diploid with additional mutation” since they are likely to
1576 harbor additional mutation(s) due to their increased fitness.

1577

1578 **Calculation of Weighted Average Z Score**

1579 To partition environments into subtle and strong perturbations of the EC, we relied on the 9
1580 experiments performed in the EC. Since each of these experiments was performed at a different
1581 time, variation in fitness across these experiments represents batch effects, and we therefore
1582 refer to these 9 experiments as “EC batches”. Environmental differences between batches are
1583 very subtle, as they represent the limit of our ability to minimize environmental variation. Thus,
1584 variation in fitness across the EC batches serves as a natural benchmark for the strength of
1585 environmental perturbations. If the deviations in fitness caused by an environmental
1586 perturbation are substantially stronger than those observed across the EC batches, we call that
1587 perturbation “strong”.

1588

1589 More explicitly, to determine whether a given environmental perturbation is subtle or strong, we
1590 first quantified the typical variation in fitness for each mutant, across the EC batches:

1591
$$\sigma_i = \frac{1}{n_{batches}} \sum_j^{batches} |f_{ij} - \bar{f}_i|$$

1592 where σ_i^2 represents the variance in fitness across the EC batches for mutant i , and
1593 \bar{f}_i represents the average fitness of mutant i across the EC batches.

1594

1595 To ensure that each mutation type contributes equally to our classification of how different each
1596 environment is from the evolution condition, we weighed each mutant’s contribution to this
1597 difference. We did so based on the number of mutants with the same mutation type, such that
1598 the mutation-type-weighted average Z-score for a given environment j is given by:

1599
$$z_j = \sum_i^{mutants} \frac{|f_{ij} - \bar{f}_i|}{n_{type(i)} \sigma_i}$$

1600 where $n_{type(i)}$ represents the number of mutants that are the same mutation type as mutant i .

1601

1602 We then classified the environmental perturbations based on this Z-score. Sixteen environments
1603 provoked fitness differences resulting in a Z-score of less than two, and we classified these
1604 environmental perturbations as “subtle”. The remaining 20 environments had Z-scores greater
1605 than 2, which we classified as “strong” environmental perturbations.

1606

1607 **Model of phenotypes that contribute to fitness**

1608

1609 In order to count the phenotypes that affect fitness in our collection of mutants, we explored a
1610 low-dimensional phenotypic model. We explicitly used a model of fitness-relevant phenotypes
1611 such that each mutant is represented as having a fixed effect on each phenotype, represented
1612 by a vector of k phenotypes, e.g. mutant i is represented by the vector $(p_{i1}, p_{i2}, p_{i3}, \dots, p_{ik})$. In
1613 addition, each environment is represented by a vector of phenotypic weights, representing the
1614 importance of each of the k phenotypes to fitness in that environment, e.g. environment j
1615 represented by the column vector $(e_{1j}, e_{2j}, e_{3j}, \dots, e_{kj})$. The fitness effect of mutant i in a given
1616 environment j is the linear combination of that mutant's phenotypes, each weighted by its
1617 importance in environment j :

$$1618 \quad f_{ij} = p_{i1}e_{1j} + p_{i2}e_{2j} + p_{i3}e_{3j} + \dots + p_{ik}e_{kj}$$

1619 Our fitness measurements reflect mutant fitness relative to a reference strain, therefore, our
1620 model places the reference strain (which has fitness 0 by definition) at the origin of this multi-
1621 dimensional space. Our model only includes phenotypes that differ between the reference strain
1622 and at least one mutant. This is sensible given that our reference strain is a modified version of the
1623 ancestor of all of these mutant lineages. Thus, if there exists a phenotype that contributes to
1624 fitness, but none of the adaptive mutants altered that phenotype, our model will not detect it.
1625 More explicitly, a phenotype that contributes to fitness would have a non-zero value of e , but if
1626 no mutant alters that phenotype from the reference, all mutants would have a zero value of p
1627 for that phenotype. Thus, the non-zero value of e would always be multiplied by a zero value for p
1628 and this phenotypic dimension would not be represented in our model. This is not to say that if
1629 only a single mutant of the 292 alters a particular phenotype we would include it as a phenotypic
1630 dimension. Our power to add dimensions to our model is limited by measurement noise. We
1631 only include dimensions that capture more variation in fitness than do dimensions that capture
1632 measurement noise (see *Estimating the detection threshold using measurement error*).

1633
1634 Similarly, because we measure fitness, and not phenotype, our model is blind to any phenotypic
1635 effect that does not contribute to fitness in at least one of the 45 environments we studied. If a
1636 mutant has large phenotypic effects, but they do not cause that mutant's fitness to differ from
1637 the reference strain in any of these 45 environments, this phenotypic effect will not be
1638 represented in our low-dimensional phenotypic model. More explicitly, mutants may have non-
1639 zero phenotypic effects p , but if these do not influence their fitness in any environment we study,
1640 e will be zero for all 45 environments. Thus, p times e will also be zero and we will not include
1641 this phenotypic dimension in our model.

1642
1643 Importantly, the phenotypic dimensions that we infer from our fitness measurements are
1644 abstract entities. They represent causal effects on fitness, rather than measurable features of
1645 cells. For this reason, they might be called "fitnotypes" (a mash of the terms "fitness" and
1646 "phenotype"). Even though the fitnotypes are independent with respect to their contribution of
1647 fitness, and contribute to fitness linearly, the mapping of commonly measured features of cells
1648 (e.g. growth rate, the expression levels of growth supporting proteins like ribosomes) onto
1649 fitnotypes may be more complicated. For instance, a commonly measured cellular feature that
1650 has a complicated nonlinear mapping to fitness could be detected as many, linearly-contributing
1651 fitnotypes. This is another reason that our phenotypic dimensions are not necessarily
1652 comparable to what people traditionally think of as a "phenotype".

1653 1654 **Using Singular Value Decomposition to decompose the fitness matrix**

1655 Our goal is to use fitness measurements to learn about the phenotypic effects of mutations as
1656 well as the contribution of these phenotypes to fitness in different environments. We conducted
1657 fitness measurements for 292 mutants in each of 45 environments and organized these data
1658 into a fitness matrix, F , where every row corresponds to a mutant, every column corresponds to
1659 an environment, and every entry is a fitness measurement. Because our model (see *Model of*
1660 *phenotypes that contribute to fitness*) represents fitness in a given environment as the sum of

1661 multiple phenotypes, each scaled by their contribution to fitness in that environment, we can use
1662 Singular Value Decomposition (SVD) to decompose the fitness matrix F as:

$$1663 \quad P\Sigma E^T = F$$

1664 The left hand side of this equation consists of three matrices: P , which represents the positions
1665 of the mutants in our low-dimensional model of phenotypic space, E^T , which represents the
1666 contribution of a phenotype to fitness in a given environment, and Σ , a diagonal matrix
1667 representing the singular values of the fitness matrix F . Though the singular values are
1668 informative in this separation of three matrices, particularly for the amount of variation captured
1669 by each of the inferred components, we can also think of this as a decomposition into two
1670 matrices, where we fold the singular values into either the mutant phenotypes or the
1671 environment weights, as described in the main text. Either way, this decomposition captures the
1672 data represented in the fitness matrix F , including measurement error as well as the underlying
1673 biological signals.

1674
1675 Importantly, the dimensions in the phenotypic model we built using SVD are detected in the
1676 order of their explanatory power. Moreover, the first dimension is the best, linear 1-component
1677 model that explains the data (if evaluated by mean squared error). This is true for any set of the
1678 first k components. This means, for example, that the model with the first eight components is
1679 the best possible 8-component linear model for the observed data (Eckart and Young, 1936).

1680
1681 One issue in this type of analysis is that adding more components always improves the
1682 explanatory power of the model, even when those components capture variation that is primarily
1683 due to measurement noise. This type of overfitting problem is common in statistics, and several
1684 methods have been devised to select the appropriate number of components to include. We use
1685 two such methods here.

1686 1687 **Estimating the detection threshold using measurement error**

1688 One method to select the appropriate number of components to include in the model and
1689 prevent overfitting (*i.e.* prevent fitting a component that primarily represents noise) is to use
1690 measurement error as a type of control. This is only possible if the amount of measurement
1691 error is known. We estimated the amount of noise in our fitness measurements using a
1692 previously described noise model (see *Noise Model*) (Venkataram et al., 2016). Since this noise
1693 model includes counting noise, every fitness measurement may have a different amount of
1694 noise. For example, mutants present at low frequency will be subject to more stochasticity
1695 resulting from counting noise. We used this noise model to simulate fitness tables (F) where
1696 mutant fitnesses vary exclusively due to measurement noise. We simulated 1000 noise-only
1697 matrices, where each entry is pulled from a normal distribution centered at zero and with
1698 variance equal to the estimated measurement noise of the corresponding entry in the true
1699 fitness matrix F . We then applied SVD to each noise-only matrix, which gave us a set of singular
1700 values generated only by noise. From many such simulations, we took the average size of the
1701 largest component, which reveals how much variation can be explained by a component that
1702 captures only noise. We found that the largest noise-components are of the size that they would
1703 capture 0.07% of variation in our true fitness matrix. Thus, we set this as our limit of detection.
1704 In other words, in order for us to include 8 components in our low-dimensional model, all of
1705 them must explain more than 0.07%% of the variation in fitness. This approach is analogous to
1706 identifying a threshold when measurement noise is known but not identical for all entries in the
1707 matrix (Josse and Sardy, 2014).

1708 1709 **Estimating detection threshold using bi-cross-validation**

1710 Another method for identifying the appropriate number of components is to use their predictive
1711 power. This method relies on the intuition that measurement error is uncorrelated across
1712 different mutants and different environments. Therefore, a component that represents
1713 measurement error should not contain information that can help predict the fitnesses of these

1714 mutants in new environments. It should also not contain information that can help predict the
1715 fitness of unstudied mutants. We used a bi-cross-validation scheme of the SVD devised by
1716 Owen and Perry (2009) which divides the mutants and environments into distinct groups of
1717 training and testing sets. This subsequently divided our matrix of fitness measurements into 4
1718 submatrices: the fitness of the training mutants in the training environments (D), the fitness of
1719 the training mutants in the testing environments (C), the fitness of the testing mutants in the
1720 training environments (B), and the fitness of the testing mutants in the testing environments (A).
1721

$$1722 \quad F = \begin{pmatrix} A & \text{Test Mutants} & B & \text{Test Mutants} \\ & \text{Test Environments} & & \text{Train Environments} \\ C & \text{Train Mutants} & D & \text{Train Mutants} \\ & \text{Test Environments} & & \text{Train Environments} \end{pmatrix}$$

1723 We carried out SVD on the training data (submatrix D), which returned a set of singular values
1724 and corresponding components that captured the fitness data in D . We then used these
1725 components to predict the fitness of the testing mutants in the testing environments (submatrix
1726 A). First, we tried to predict these fitness values by only using the first component. That is, we
1727 fixed this first component and the first singular value for the training mutants. We then found the
1728 best first component for the testing environments based on the fitness values of the training
1729 mutants in these environments (i.e. using the information in submatrix D), given the constraint
1730 that the training mutants can only be represented by the one component. We then conducted an
1731 analogous procedure to find the first component of the testing mutants by fixing the first
1732 component of the training environments by using the information in submatrix B . Then, we tried
1733 to predict the fitness of the testing mutants in the testing environments using the first component
1734 independently fit for each. We subsequently repeated this procedure, giving the testing mutants
1735 access to more of the training components each time. If the components detected by the
1736 training components represent biological signal, then this should improve the ability to predict
1737 the fitness of the testing mutants in the testing environments. However, once the components
1738 primarily represent measurement error, their inclusion should harm predictive power. Therefore,
1739 we use the number of components with the best ability to predict the held-out data (submatrix A)
1740 as the number of components that represent biological signal in our data.
1741

1742
1743 For computational efficiency, we explicitly used the formulation proposed by Owen and Perry
1744 (2009) for the prediction of the held-out submatrix A :

$$1745 \quad \hat{A} = B(\hat{D}^{(k)})^+ C$$

1746 where $(\hat{D}^{(k)})^+$ denotes the Moore-Penrose inverse of the rank k approximation of sub-matrix D .
1747 This prediction is equivalent to the procedure outlined above, provided that least-squares
1748 regression is used to identify the components of the testing mutants and testing conditions,
1749 conditional upon the training components (Owen and Perry, 2009).
1750

1751 We divided our mutants into fixed training and testing sets (see *Division of Mutants into Training
1752 and Testing Sets*) and used these sets throughout our study. As for training versus testing
1753 environments, these changed depending on our goal. For validating the number of components
1754 to include in our phenotypic model, we held out each of the 25 subtle environmental
1755 perturbations, using it as the testing environment and the other 24 for training. For making
1756 predictions of the fitness of the testing mutants in the strong environmental perturbations, we
1757 used all 25 subtle environmental perturbations as the training set, though we also show how
1758 these predictions vary when each of the 25 subtle environmental perturbations is held out from
1759 the training set.
1760

1761 **Division of Mutants into Training and Testing Sets**

1762 In order to perform bi-cross-validation on our data, we need to divide our data into training and
1763 testing sets. Because some mutation types, in particular diploids and Ras/PKA mutants, are
1764 present more than others in our collection of mutants, we sampled the training set such that

1765 each mutation type is represented roughly equally (see *Classifying mutants by mutation type*).
1766 Specifically, we designated half of each mutation type, with a maximum of 20 representatives of
1767 each type, as belonging to the training set. The remaining mutants comprise the test set. For
1768 example, there are 188 diploids included in the 292 adaptive mutants. We included 20 in the
1769 training set and 168 in the test set. There are 20 *IRA1-nonsense* mutants included in the 292,
1770 and we included 10 in the training and 10 in the test set. Additionally, genes that are
1771 represented only once in the set of mutations are placed in the test set. This results in a training
1772 set of 60 mutants and a testing set of 232 mutants (see Table S1).

1773

1774 **Using simulated data to validate detection threshold estimation**

1775 To further validate our approach for identifying the number of detectable phenotypic
1776 components from our data, we simulate data that consists of a known number of phenotypic
1777 components k and use our methods to estimate the number of phenotypic components
1778 detectable in the data. To simulate the phenotype space, we place 100 mutants at random in
1779 the k -dimensional phenotype space P . The coordinates of these mutants are pulled from a
1780 uniform distribution in the n -ball (e.g. the n -ball is a sphere if there are three dimensions)
1781 centered at coordinates $(1, 0, \dots, 0)$ with radius 1. We center the mutants at 1 in the first
1782 dimension and 0 in all other dimensions in order to create data similar to our empirical data
1783 where the first component captures much of the variation in fitness. We then similarly place 50
1784 environments at random in the k -dimensional environmental space E . Recall that this space
1785 represents the importance of each phenotype in each environment (see **Fig 3** and see *Model of*
1786 *phenotypes that contribute to fitness*). The environments are pulled from a uniform distribution in
1787 the n -ball centered at $(1, 1, \dots, 1)$ with a small radius of 0.1 chosen such that the environmental
1788 perturbations are subtle. Note, for computational efficiency, we use the algorithm from
1789 (Marsaglia, 1972) to pull points uniformly distributed in the n -ball. Next, we calculate the fitness
1790 of each of these mutants in each environment as a linear combination of the mutant's
1791 phenotypes weighted by the contribution of each of these phenotypes to fitness in the relevant
1792 environment (see *Model of phenotypes that contribute to fitness*). We then add measurement
1793 error to these fitness values to simulate the effect that measurement uncertainty has on our
1794 ability to detect phenotypic components. We simulate the data with various numbers of
1795 phenotypic components (2, 3, 4, 5, 10, 20, 30, 40, and 49) and use our methods to try to
1796 estimate the number in each set.

1797

1798 We find that our method for identifying the number of detectable phenotypic components from
1799 the measurement error (see *Estimating the detection threshold using measurement error*)
1800 accurately identifies the simulated number of phenotypic components when measurement noise
1801 is very low (Fig S1B). As measurement noise increases, our approach detects fewer
1802 components, as expected due to measurement noise swamping the smallest components of
1803 signal (Fig S1B). Bi-cross-validation, which holds out each environment and half of the mutants
1804 (see *Estimating the detection threshold using bi-cross-validation*), performs similarly, detecting
1805 the appropriate number of phenotypic components when measurements are sufficiently precise
1806 (Fig S1C).

1807

1808 **Clustering mutants in phenotype space**

1809 After inferring the low-dimensional model of phenotype space using SVD, we used Uniform
1810 Manifold Approximation and Projection (UMAP) to visualize how the mutants cluster in that
1811 space. For this analysis, we used the 8-component phenotypic model that we built from the 60
1812 training mutants and the 25 subtle perturbations. We did this to avoid the model being
1813 dominated by variation in very common mutations, specifically the diploids, which make up
1814 188/292 of our adaptive mutants. We added more mutants in the visualization by finding the
1815 location of each of the testing mutants (except diploids) by least sum of squares optimization.
1816 To do so we fixed the coordinates for the 25 environments and found the coordinates for each
1817 mutant that best estimated its fitness in all environments. To further avoid our visualization

1818 being dominated by the diploids, we included only the diploids present in the training set in our
1819 visualization. For UMAP, we specified that 20 neighbors are used.

1820
1821 Though UMAP tends to preserve both local and global structure (McInnes et al., 2018) it is not
1822 necessarily representative of the distance between objects in high-dimensional space. Thus, to
1823 quantify more precisely the clustering by gene observed, we explicitly compared the median
1824 pairwise distance between these apparent clusters to 10000 randomly chosen sets of the same
1825 size and calculated empirical p-values. Because there are many diploids such that they will be
1826 the most prevalent type of mutant drawn in these randomly chosen sets, we only drew from
1827 strains that have other mutations besides or in addition to diploidy. We use the median pairwise
1828 distance, rather than the mean, to identify the typical distance between mutants in a given
1829 cluster to reduce the influence outlier mutations that might bias the mean pairwise distance.

1830 1831 **Calculation of Weighted Coefficient of Determination**

1832 Because mutants are present in unequal numbers in the test set, standard measures of
1833 variance explained are likely to be representative of our ability to predict mutants that have
1834 many barcoded lineages present in the data, for instance diploid and IRA1-nonsense mutations.
1835 These measures would be less representative of mutants with few lineages present, i.e.
1836 TOR/Sch9 pathway mutants. Thus, we use a measure of predictability (\tilde{R}^2) that weights the
1837 contribution of each mutant to overall variance explained based on the number of lineages that
1838 share its mutation type (diploids, IRA1 nonsense, IRA1 missense, GPB2, etc.). This effectively
1839 measures our ability to predict the fitness of each mutation type, rather than each mutant. . For
1840 overall predictive power across all mutants and conditions, we used the measure:

$$1841 \quad \tilde{R}^2 = 1 - \frac{\sum_i^{\text{mutants}} \sum_j^{\text{conditions}} \frac{1}{n_{\text{type}(i)}} (f_{ij} - \widehat{f}_{ij})^2}{\sum_i^{\text{mutants}} \sum_j^{\text{conditions}} \frac{1}{n_{\text{type}(i)}} (f_{ij} - \bar{f})^2}$$

1842 where \bar{f} denotes the average fitness for all evaluated mutants and evaluated conditions.

1843 We used a similar measure to quantify the ability to predict fitness for each environment j . This
1844 is given by:

$$1845 \quad \tilde{R}_j^2 = 1 - \frac{\sum_i^{\text{mutants}} \frac{1}{n_{\text{type}(i)}} (f_{ij} - \widehat{f}_{ij})^2}{\sum_i^{\text{mutants}} \frac{1}{n_{\text{type}(i)}} (f_{ij} - \bar{f}_j)^2}$$

1846 where \bar{f}_j denotes the average fitness across all evaluated mutants in condition j .

1847
1848 Note that this measure explicitly compares a model's fitness prediction in each environment to
1849 predictions made using the average fitness in that environment, such that if the model's fitness
1850 prediction is the same as the average fitness, \tilde{R}^2 is zero. It is possible that a given model's
1851 fitness prediction is worse than that of the average fitness in that environment, resulting in
1852 negative values of \tilde{R}^2 . In our work, negative \tilde{R}^2 values occur for the 1-component model when
1853 predicting the fitness of mutants in some of the strong environmental perturbations. In particular,
1854 this occurs when fitness in that environment is uncorrelated with EC fitness, which is captured
1855 by the first component, such that the EC fitness is unable to make reasonable predictions of
1856 fitness in this environment.

1857
1858 Note that we observe qualitatively similar results to this measure when we use a standard
1859 variance explained measure and exclude diploids, which dominate the test set (see Fig S5).

1860
1861
1862
1863

1864 **Calculating mutant-specific improvement**

1865

1866 It is possible that all 292 of our adaptive mutants each affect all 8 of the phenotypic components
1867 in our low-dimensional model, however, it is also possible that some mutants influence some
1868 phenotypes more strongly than others. In order to quantify how much a specific component
1869 lends to the ability to predict the fitness of each mutant in each environment, we need a metric
1870 to calculate the difference in predictive accuracy for the model with and without this component.
1871 Specifically, to assess the impact of the inclusion of the k th component, we compared the
1872 prediction accuracy of the k -component model to the model that includes the first $k-1$
1873 components.

1874

1875 Because fitness estimates vary in their reliability due to finite coverage and other sources (see
1876 *Noise model* section), we should factor this uncertainty in our measure of prediction
1877 improvement. For example, a small improvement in prediction accuracy for a very uncertain
1878 fitness estimate is less meaningful than the same improvement in prediction accuracy for a
1879 fitness estimate that we are quite confident in. Thus, we scale the difference in prediction
1880 accuracy by the amount of uncertainty in the underlying fitness estimate.

1881

1882 This gives us the measure of improvement in the estimate of the fitness of mutant i in condition j
1883 due to the inclusion of the n th component as:

1884

$$I_{ij}^k = \frac{(\widehat{f}_{ij}^{k-1} - f_{ij}) - (\widehat{f}_{ij}^k - f_{ij})}{\epsilon_{ij}}$$

1885 where \widehat{f}_{ij}^k and \widehat{f}_{ij}^{k-1} represent the estimate of the fitness of mutant i in condition j for the model
1886 with k and $k-1$ components, respectively. f_{ij} and ϵ_{ij} represent the measured fitness value and
1887 measurement uncertainty for the fitness of mutant i in condition j , respectively.

1888

1889 **DATA AND CODE AVAILABILITY**

1890

1891 **Data Resource**

1892 The raw Illumina sequencing data for the fitness measurement assays conducted in this study
1893 can be found under NIH BioProject: PRJNA641718. Sequencing data previously published in
1894 Venkataram et al., 2016 can be found under NIH BioProject: PRJNA310010. Sequencing data
1895 previously published in Li et al., 2018 can be found under NIH BioProject: PRJNA388215.

1896

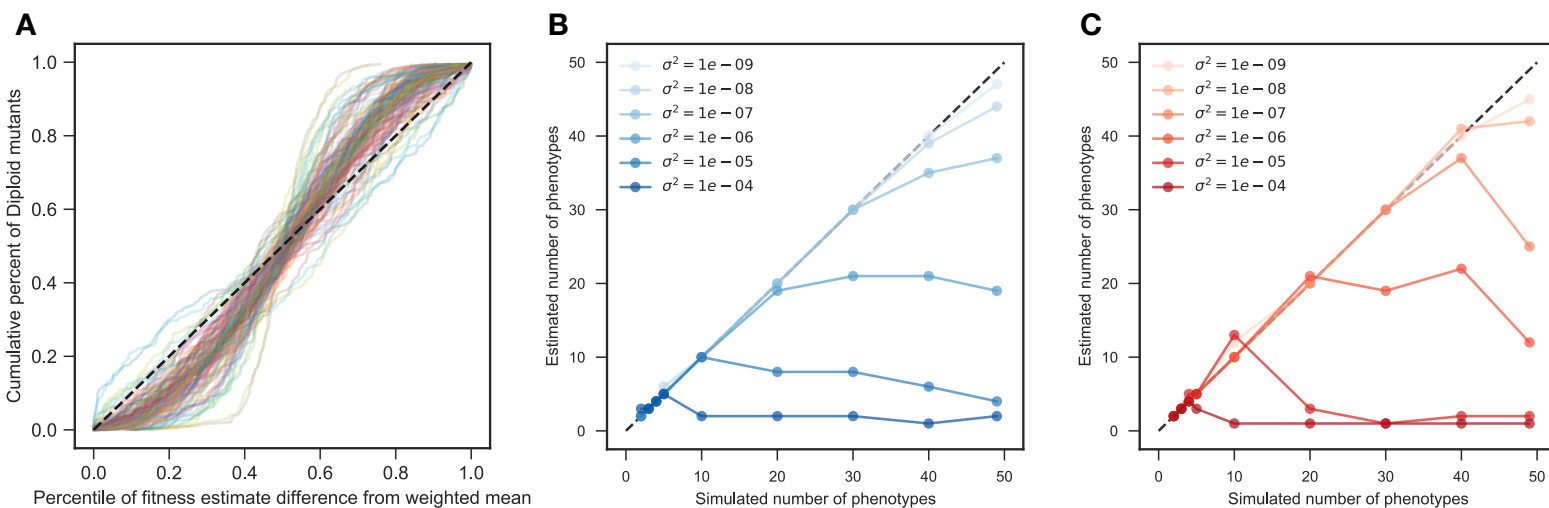
1897 **Code**

1898 The software repository for the barcode counting code can be found at
1899 <https://github.com/sandeepvenkataram/BarcodeCounter2>.

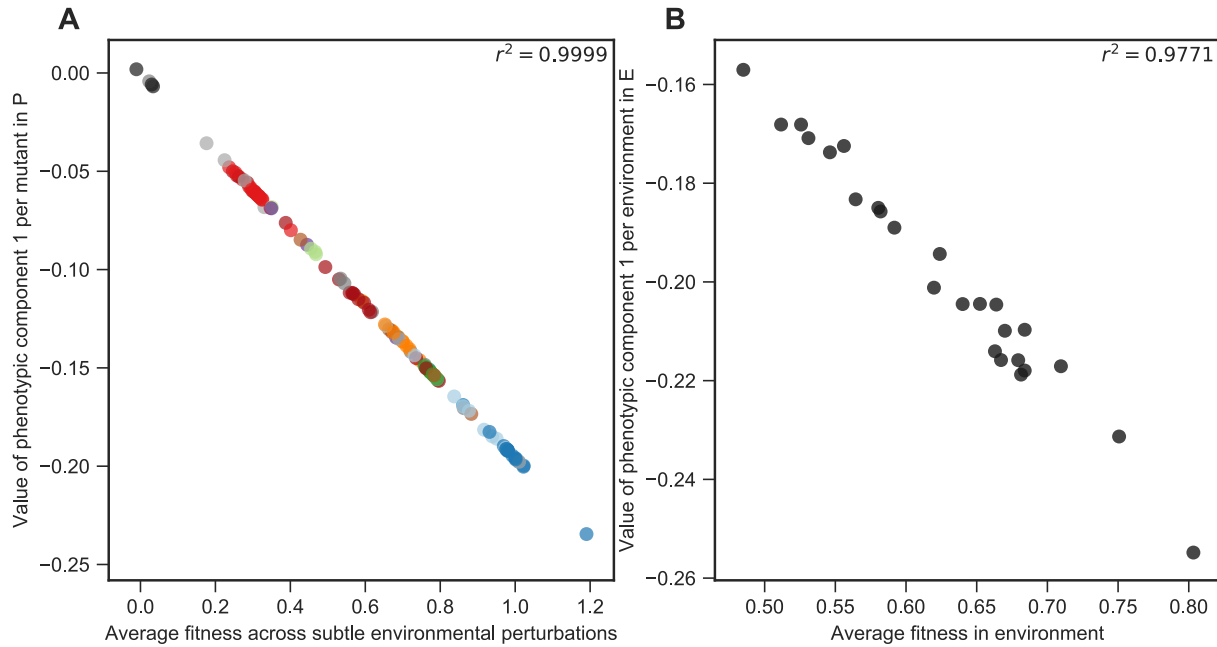
1900 The software repository for the fitness estimate inference can be found at
1901 <https://github.com/barcoding-bfa/fitness-assay-python>.

1902 The code for all downstream analysis, including figure generation can be found at
1903 <https://github.com/grantkinsler/1BigBatch>.

1904 SUPPLEMENTARY FIGURES

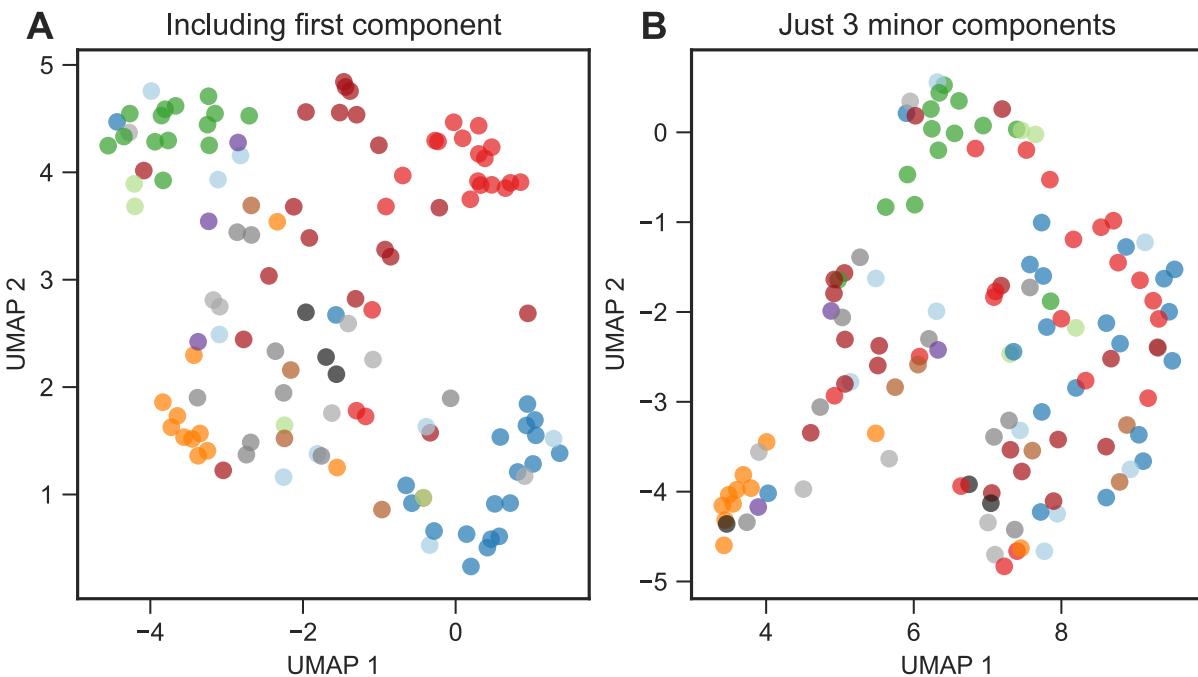


1905 **Fig S1. Noise model is a conservative measure of uncertainty.** (A) Fitness differences among strains
 1906 that are genetically identical and have very similar fitness effects tell us about the amount of
 1907 measurement noise. Our strain collection includes 188 diploids that have similar fitnesses and possess
 1908 no mutations other than diploidy. For each diploid fitness estimate, we calculated the percentile of
 1909 deviation from the weighted average of all diploid fitness estimates in a particular environment. This is
 1910 shown on the horizontal axis. The vertical axis shows the cumulative percent of diploids with deviations
 1911 listed on the horizontal axis. If the noise model perfectly captures the uncertainty of each measurement,
 1912 then it should be represented by the black dashed line, as, for instance, 20% of the diploids should have
 1913 a difference from the mean in the 20th percentile. Each line represents a single experiment (we have 45
 1914 environments each with several replicates for a total of 109 experiments, see Methods). For the vast
 1915 majority of experiments, the diploids are closer to the mean than predicted by our noise model, as
 1916 indicated by each line's sigmoidal shape. This indicates that the noise model is conservative. (B) The
 1917 horizontal axis represents the number of phenotypic components in simulated data consisting of 100
 1918 mutants and 50 environments. The vertical axis indicates the number of components we detected when
 1919 we only count components that explain more variation than does our noise model (see Methods). For low
 1920 levels of measurement noise (light blue), our method accurately detects the number of simulated
 1921 components. As measurement noise increases (darker blue dots), the noise begins to swamp signal and
 1922 the number of detected components decreases. (C) Same as (B), but when we set the threshold for
 1923 detecting components using bi-cross validation rather than our noise model by holding out each
 1924 environment and half of the mutants. Darker color indicates more measurement noise.
 1925



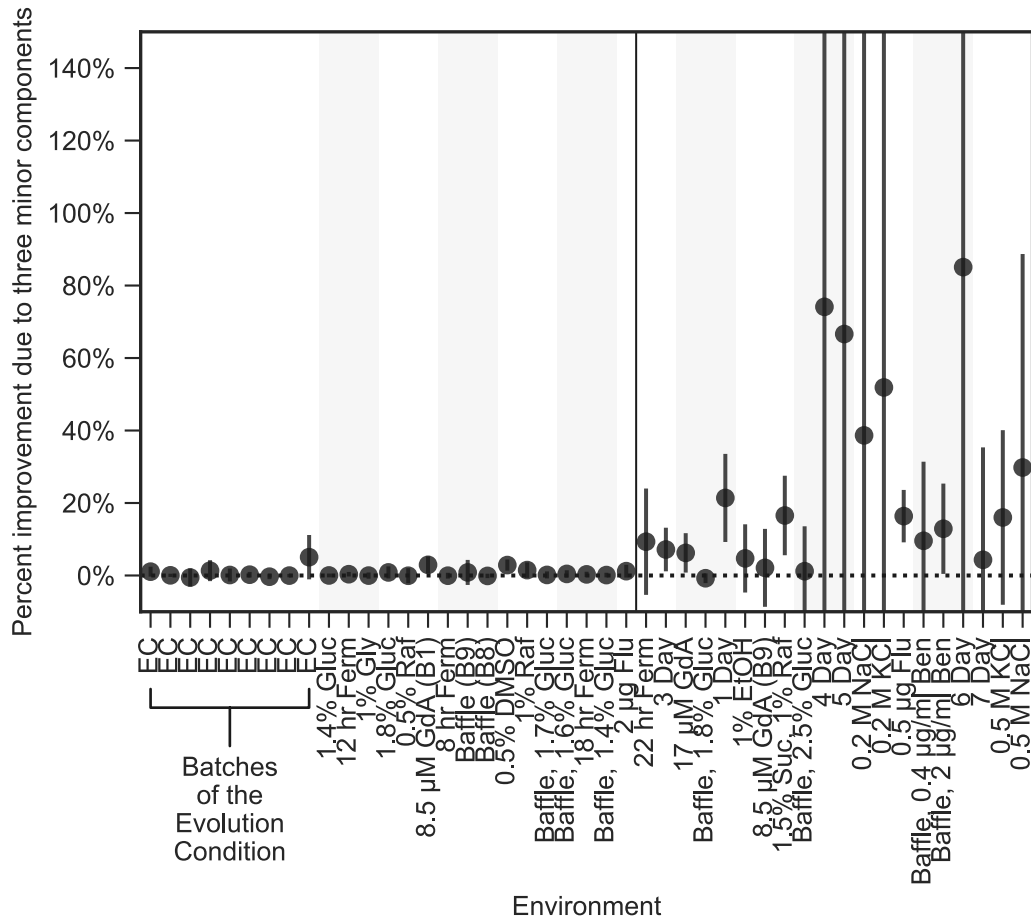
1926
1927
1928
1929
1930
1931
1932
1933
1934

Figure S2. The first component represents the mean fitness of each mutant in the 25 subtle perturbations, as well as the mean impact of each perturbation on fitness. (A) The horizontal axis shows the average fitness of each mutant across all 25 environments that represent subtle perturbations. The vertical axis shows the value of the first phenotypic component for each mutant. Mutants are colored as in Figure 2. **(B)** The horizontal axis shows the average fitness of all 292 mutants in each environment. The vertical axis shows the value of the first phenotypic component in the environment weight space *E*.



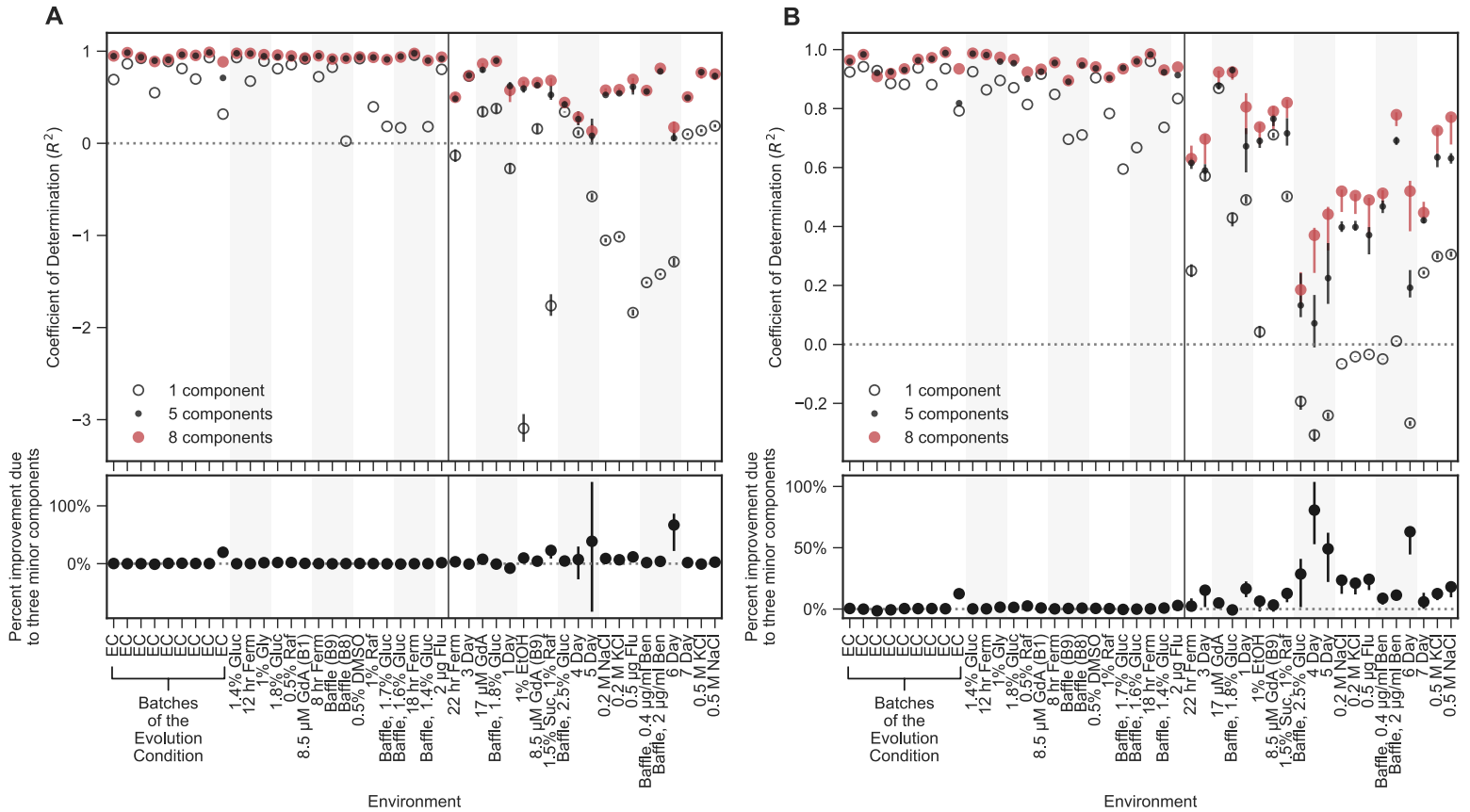
1935
1936
1937
1938
1939
1940
1941

Fig S3. Low-dimensional phenotypic models, and subsets of such models, cluster mutants by gene and mutation type. (A) UMAP clusters mutants visually by gene when using the full 8-component phenotype space. **(B)** UMAP also shows some clustering when using only the three components that explain the least variation in mutant fitness. Though the clustering is clear for *PDE2* and *GPB2*, it less clearly delineates *IRA1-nonsense* and diploid mutants. This suggests these mutants do not have substantial effects on these phenotypic components.



1942
1943
1944
1945
1946
1947

Fig S4. Improved fitness predictions when including the three smallest phenotypic components is not specific to choice of training mutants. This plot is similar to the lower panel of figure 4A, except here, black dots indicate the average improvement across 100 choices of the training and test sets. Error bars indicate two standard deviations from the mean.



1948 **Fig S5. Prediction ability using unweighted coefficient of determination.** These plots are similar to
 1949 figure 4A except here the vertical axis displays prediction power using a standard, rather than a weighted,
 1950 coefficient of determination measure. Because diploids dominate the number of mutants in the collection,
 1951 there are large differences between panel A (which shows all mutants) and panel B (which omits
 1952 diploids).

1953 **SUPPLEMENTARY TABLES**

Mutation Type	Total Number	Number in Training Set	Number in Test Set
Diploid	188	20	168
Diploid w/ add'l mutation			
Diploid + Chr11Amp	3	1	2
Diploid + Chr12Amp	1	0	1
Diploid + IRA1	1	0	1
Diploid + IRA2	3	1	2
High-fitness Diploid	11	5	6
IRA1			
IRA1 nonsense	20	10	10
IRA1 missense	9	4	5
IRA1 other	1	0	1
IRA2	8	4	4
GPB1	4	2	2
GPB2	14	7	7
PDE2	11	5	6
Other Ras/PKA pathway			
CYR1	3	1	2
TFS1	1	0	1
RAS2	1	0	1
TOR/Sch9 pathway			
KOG1	1	0	1
SCH9	1	0	1
TOR1	1	0	1
Other Adaptive	7	0	7
Neutral	3	0	3
TOTAL	292	60	232

Table S1. List of all mutants included in this study.

1954
1955
1956

#	Environment Name	Source	Number of barcoded clones	Batch Number	Number of replicates	Description of Manipulation
1	EC	This study	500	10	3	No preculture
2	EC	This study	4800	4	3	No barcodeless ancestor
3	EC	Li et al (2018)	4800	4	3	
4	EC	This study	500	5	3	No preculture
5	EC	Venkataram et al (2016)	4800	1	3	
6	EC	Li et al (2018)	4800	3	3	
7	EC	Venkataram et al (2016)	4800	2	3	
8	EC	This study	500	9	4	
9	EC	This study	500	5	3	
10	1.4% Gluc	This study	500	9	2	1.4% glucose concentration
11	12hr Ferm	Li et al (2018)	4800	3	3	44 hours of growth, 2×10^8 cells transferred, resulting in ~12 hours of fermentation phase
12	1% Gly	This study	500	9	2	Added 1% glycerol
13	1.8% Gluc	This study	500	9	2	1.8% glucose concentration
14	0.5% Raf	This study	500	9	2	Added 0.5% raffinose
15	8.5 uM GdA (B1)	This study	4800	1	3	Added 8.5uM Geldanamycin
16	8hr Ferm	Li et al (2018)	4800	3	3	40 hours of growth, 8×10^8 cells transferred, resulting in ~8 hours of fermentation phase
17	Baffle (B9)	This study	500	9	2	Used baffled flask
18	Baffle (B8)	This study	4800	8	2	Used baffled flask
19	0.5% DMSO	This study	4800	1	3	Included 0.5% DMSO
20	1% Raf	This study	500	9	2	Added 1% raffinose
21	Baffle, 1.7% Gluc	This study	4800	8	2	1.7% glucose concentration, used baffled flask
22	Baffle, 1.6% Gluc	This study	4800	8	2	1.6% glucose concentration, used baffled flask
23	18hr Ferm	Li et al (2018)	4800	4	3	50 hours of growth, 2.5×10^7 cells transferred, resulting in ~18 hours of fermentation phase
24	Baffle, 1.4% Gluc	This study	4800	8	2	1.4% glucose concentration, used baffled flask
25	2 ug Flu	This study	500	9	2	Added 2ug Fluconazole
26	22hr Ferm	Li et al (2018)	4800	4	3	54 hours of growth, 6.25×10^6 cells transferred, resulting in ~22 hours of fermentation phase
27	3 Day	Li et al (2018)	4800	2	3	3 days of growth
28	17uM GdA	This study	500	9	2	Added 17uM Geldanamycin
29	Baffle, 1.8% Gluc	This study	4800	8	2	1.8% glucose concentration, used baffled flask
30	1 Day	Li et al (2018)	4800	2	3	24 hours of growth
31	1% EtOH	This study	500	9	2	Added 1% ethanol
32	8.5uM GdA (B9)	This study	500	9	2	Added 8.5uM Geldanamycin
33	1.5% Suc, 1% Raf	This study	500	9	2	No glucose, 1.5% Sucrose, 1% Raffinose
34	Baffle, 2.5% Gluc	This study	4800	8	2	2.5% glucose concentration, used baffled flask
35	4 Day	Li et al (2018)	4800	2	3	4 days of growth
36	5 Day	Li et al (2018)	4800	6	3	5 days of growth
37	0.2 M NaCl	This study	500	9	2	Added 0.2M NaCl
38	0.2 M KCl	This study	500	9	1	Added 0.2M KCl
39	0.5 ug Flu	This study	500	9	2	Added 0.5ug Fluconazole
40	Baffle, 0.4 ug/ml Ben	This study	4800	7	2	Added 0.4ug/mL Benomyl, used baffled flask
41	Baffle, 2 ug/ml Ben	This study	4800	7	2	Added 2ug/mL Benomyl, used baffled flask
42	6 Day	Li et al (2018)	4800	6	3	6 days of growth
43	7 Day	Li et al (2018)	4800	6	3	7 days of growth
44	0.5 M KCl	This study	500	9	1	Added 0.5M KCl
45	0.5 M NaCl	This study	500	9	1*	Added 0.5M NaCl

1957 **Table S2.** List of all conditions used in this study, ordered by deviation from the EC batch as in the main
1958 text figures.