

Quantifying the Adaptive Potential of a Nascent Bacterial Community

Joao A Ascensao¹, Kelly M Wetmore², Benjamin H Good³, Adam P Arkin^{1,2}, and Oskar Hallatschek⁴

¹Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA

²Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA

³Department of Applied Physics, Stanford University, Stanford, CA, USA

⁴Departments of Physics and Integrative Biology, University of California Berkeley, Berkeley, CA, USA

1 The fitness effects of all possible mutations available to an organism largely shapes the dynamics of evolutionary adaptation. Tremendous progress has been made in quantifying the strength and abundance of selected mutations available to single microbial species in simple environments, lacking strong ecological interactions. However, the adaptive potential of strains that are part of multi-strain communities remains largely unclear. We sought to fill this gap for a stable community of two closely related ecotypes (“L” and “S”) shortly after they emerged within the *E. coli* Long-Term Evolution Experiment (LTEE). To this end, we engineered genome-wide barcoded transposon libraries and developed a computational inference pipeline to measure the fitness effects of all possible gene knockouts in the coexisting strains as well as their ancestor, for many different conditions. We found that the fitness effect of most gene knockouts sensitively depends on the genetic background and the ecological conditions, as set by environmental perturbations and the relative frequency of both ecotypes. Despite the idiosyncratic behavior of individual knockouts, we still see consistent statistical patterns of fitness effect variation across both genetic background and community composition. The background dependence of mutational effects appears to reflect widespread changes in which gene functions are important for determining fitness, for all but the most strongly interacting genes. Additionally, fitness effects are correlated with evolutionary outcomes for a number of conditions, possibly revealing shifting patterns of adaptation. Together, our results reveal how ecological and epistatic effects combine to drive adaptive potential in recently diverged, coexisting ecotypes.

30 Experimental evolution | Barcoded transposon mutagenesis | Distribution of fitness effects | Eco-evolutionary dynamics | Global epistasis
31
32 Correspondence: ohallats@berkeley.edu

33 Introduction

34 The fitness effects of all possible mutations largely shapes the adaptive routes available to an organism and the overall dynamics of adaptation. Population genetic theory has helped us to understand what evolutionary dynamics we should expect if we know the fitness effects of all available mutations, i.e. the "distribution of fitness effects" (DFE), and how the distribution changes due to epistasis and evolution (1–6). Additionally, knowing the selective strength of a given mutation can help to predict the chance that the mutation establishes in a population. However, high-quality measurements of the DFE in a given system requires sampling and measuring the fitness effects of sufficiently many mutations across the genome. This has only become possible recently (primar-

ily in microbes), due to the rise of sequencing technologies. DNA barcoding systems have become especially influential to better understand microbial adaptive evolution. By taking advantage of amplicon sequencing methods to measure barcode frequency dynamics, these systems have been used with great success to directly observe evolutionary dynamics (7–10), identify selected mutations and the statistical patterns that characterize them (11–13), and uncover the pleiotropic effects of mutations across environments (14–16).

However, the spectrum of mutational effects, and thus the adaptive potential, in strains that are part of multi-strain communities remains largely unclear. The notion of a single-dimensional DFE may not even apply to community members, as the community evolution may depend on how a mutation changes adaptation to all the relevant environmental niches. Thus, the coexisting ecotypes of a community are ecologically and evolutionarily coupled to each other in an eco-evolutionary process that begins as soon as a new ecotype arrives in community. The process that generates new ecotypes from existing community members, ecological diversification, was previously considered to be a relatively rare event. However, spontaneous and rapid diversification appears to be pervasive in many natural microbiomes (17–19) and experimental systems, where ecological stability is usually attained by niche partitioning (20–22) or cross-feeding (23, 24). A canonical example of these eco-evolutionary dynamics have been observed in the *E. coli* Long Term Evolution Experiment (LTEE)—an experiment that has tracked the evolution of several *E. coli* populations over the course of over 70,000 generations (at the time of writing). Through time-resolved metagenomic sequencing of the LTEE, it was found that 9 out of the 12 populations evolved two separate lineages that coexisted with each other for tens of thousands of generations, while continuing to accumulate mutations and adapt (25). Together, these observations suggest that spontaneous diversification followed by coevolution is a major adaptive route for microbial populations.

Despite the ubiquity of eco-evolutionary dynamics characterized by diversification and coevolution, we have limited theoretical understanding (26, 27) and experimental measurements of how mutational effects should depend on the (a)biotic environment, and how they drive evolutionary dynamics. Thus, we sought to (i) better understand how the spectrum of mutational effects changes over long evolutionary timescales and after ecological diversification, (ii)

92 how mutational fitness effects depend on ecological condi- 148
93 tions, and (iii) if and how they are correlated with evolu- 149
94 tionary outcomes, i.e. establishment of mutations in a gene 150
95 and changes in gene expression. To sample mutational fit- 151
96 ness effects, we created genome-wide knockout libraries via 152
97 random-barcoded transposon mutagenesis (28, 29). While 153
98 knockout mutations do not represent all possible mutations 154
99 in the genome, this approach allows us to sample a wide vari- 155
100 ety of mutations across the genome and to compare the effect 156
101 of the same mutation across different genetic backgrounds. 157
102 We used a population from the LTEE as a model system, pri- 158
103 marily because we can sample and probe clones right after 159
104 the ecological diversification event. This allows us to bet- 160
105 ter understand the adaptive potential in a nascent commu- 161
106 nity and how that then affects further coevolution. Early in 162
107 the LTEE, one population, ara-2, spontaneously diversified 163
108 into two lineages that coexist via negative frequency depen- 164
109 dence, termed S and L (for their small and large colony sizes 165
110 on certain agar plates) (30). S and L coexist by inhabiting 166
111 different temporal niches in the LTEE environment, set up 167
112 as serial dilutions in glucose minimal media—L grows more 168
113 quickly on glucose during exponential phase, while S spe- 169
114 cializes on stationary phase survival and utilizing acetate, a 170
115 byproduct of overflow metabolism (31, 32). Following di- 171
116 versification, the lineages have persisted to this day and con- 172
117 tinued to evolve and adapt, diverging on genetic, transcrip- 173
118 tional, and metabolic levels (25, 30–35). Because of the long, 174
119 documented evolutionary history of the LTEE as well as our 175
120 barcoding approach, our data allows for the quantification 176
121 of how fitness effects change due to evolution and ecology, 177
122 presenting opportunities to identify general patterns of eco- 178
123 evolutionary processes. 179

124 Results

125 **Measuring knockout fitness effects.** We sought to mea- 181
126 sure the knockout fitness effects available to the small LTEE- 182
127 derived ecosystem of S and L, and how they depend on 183
128 ecological conditions, including the abundance of its con- 184
129 stituents. To this end, we created randomly barcoded trans- 185
130 poson libraries of three LTEE clones, using previously devel- 186
131 oped methods (RB-TnSeq) (28, 29)—S and L clones sampled 187
132 from 6.5k generations, right after diversification (30, 34), and 188
133 their LTEE ancestor, REL606 (Figure 1A). We used these li- 189
134 braries to measure the knockout fitness effects of nearly all 190
135 non-essential genes in various environments relevant to the 191
136 evolution of the population in the LTEE (Table 1), by propa- 192
137 gating the libraries in defined conditions (with two biological 193
138 replicates per experiment) and using Illumina amplicon se- 194
139 quencing to track the frequency trajectories of different bar- 195
140 codes (Figure 1B). By essentially measuring the log-slope of 196
141 the frequency trajectories, we can estimate the fitness effect, 197
142 s , of a given mutant (Figure 1C), which we report in units 198
143 of $1/\text{generation}$. Transposon insertion events were highly re- 199
144 dundant, with a median of ~ 20 insertions per gene, allowing 200
145 us to combine information from multiple barcode trajec- 201
146 tories into one fitness measurement through our statistical fit- 202
147 ness inference pipeline and identify significantly non-neutral 203
204

148 mutations (FDR correction; $\alpha = 0.05$). We carefully quanti- 149
150 fied sources of error in barcode frequency measurements and 151
152 propagated them to our fitness estimates, which was crucial 153
154 to effective and accurate analysis of the data (see supplement 155
156 section S3)—for example, we could exclude knockouts with 157
158 overly noisy fitness measurements, or weight measurements 159
160 by their error. 161

162 After inferring the fitness effect of each gene knockout, we 163
164 can compare fitness effects across genetic backgrounds and 165
166 environments. We can first look at knockout fitness effects 167
168 in the evolutionary condition proxies—the closest approxima- 169
170 tion to the environment where evolution in the LTEE took 171
172 place: the REL606 library in monoculture, and S and L li- 173
174 braries together, coexisting at the ecological equilibrium fre- 174
175 quency. In coculture experiments, the S/L libraries are mixed 175
176 in the minority together with wild-type S/L clones at the de- 176
177 sired frequency (see supplement section S2). The ecotype 177
178 frequencies do not change considerably over the time period 178
179 considered (Figure S6). As previously mentioned, the overall 179
180 DFE and how it changes due to evolution largely controls the 180
181 speed of adaptation (1–6). If we look at the overall DFE in 181
182 the evolutionary condition proxies, we see that REL606 has 182
183 access to beneficial knockouts of much larger effect size than 183
184 either S or L (Figure 1D), suggesting that REL606 would 184
185 adapt much quicker than S or L. The evolutionary tendency 185
186 towards a "shrinking DFE" is known as global diminishing 186
187 returns epistasis, which has previously been proposed as a 187
188 mechanism to explain the decelerating fitness trajectories of 188
189 the LTEE populations (37, 38). While diminishing returns 189
190 epistasis was previously observed to affect the first couple 190
191 common LTEE mutations (39), global diminishing returns 191
192 (affecting the whole DFE) after the accumulation of many 192
193 mutations had not yet been directly observed. 193

194 We can also compare the fitness effects of each knockout 194
195 mutation both between replicates and across genetic back- 195
196 grounds (Figure 1E-F), to contrast within-sample to between- 196
197 sample variance. In contrast to a strong replicate-replicate 197
198 correlation, we see that fitness effects are largely uncorre- 198
199 lated between genetic backgrounds; so, it appears that in- 199
200 dividual mutations behave idiosyncratically despite statisti- 200
201 cal patterns of epistasis, in contrast with previous experi- 201
202 ments (39, 40) which saw diminishing returns both glob- 202
203 ally and with individual mutations. Most knockout muta- 203
204 tions that were strongly beneficial in REL606 and then ac- 204
205 quired a mutation in that gene in the 6.5k S/L background 205
206 became effectively neutral when knocked out in S/L (*nadR*, 206
207 *pykF*, *ybaL*, *ygaZ*); it makes sense that mutating a gene that 207
208 was already mutated (with a fitness effect) wouldn't have 208
209 an effect, if the mutation was effectively a loss-of-function. 209
210 One gene, *spoT*, was beneficial in REL606 but deleterious 210
211 in both S and L when knocked out, indicating that the nat- 211
212 ural *spoT* SNP likely represents a change-of-function rather 212
213 than a loss-of-function mutation. However, the majority of 213
214 selected genes in REL606 were not mutated between 0 and 214
215 6.5k generations in S/L, so the fact that their fitness effects 215
216 significantly changed across genetic background implicates 216
217 the role of widespread, global idiosyncratic epistasis. Fur- 217

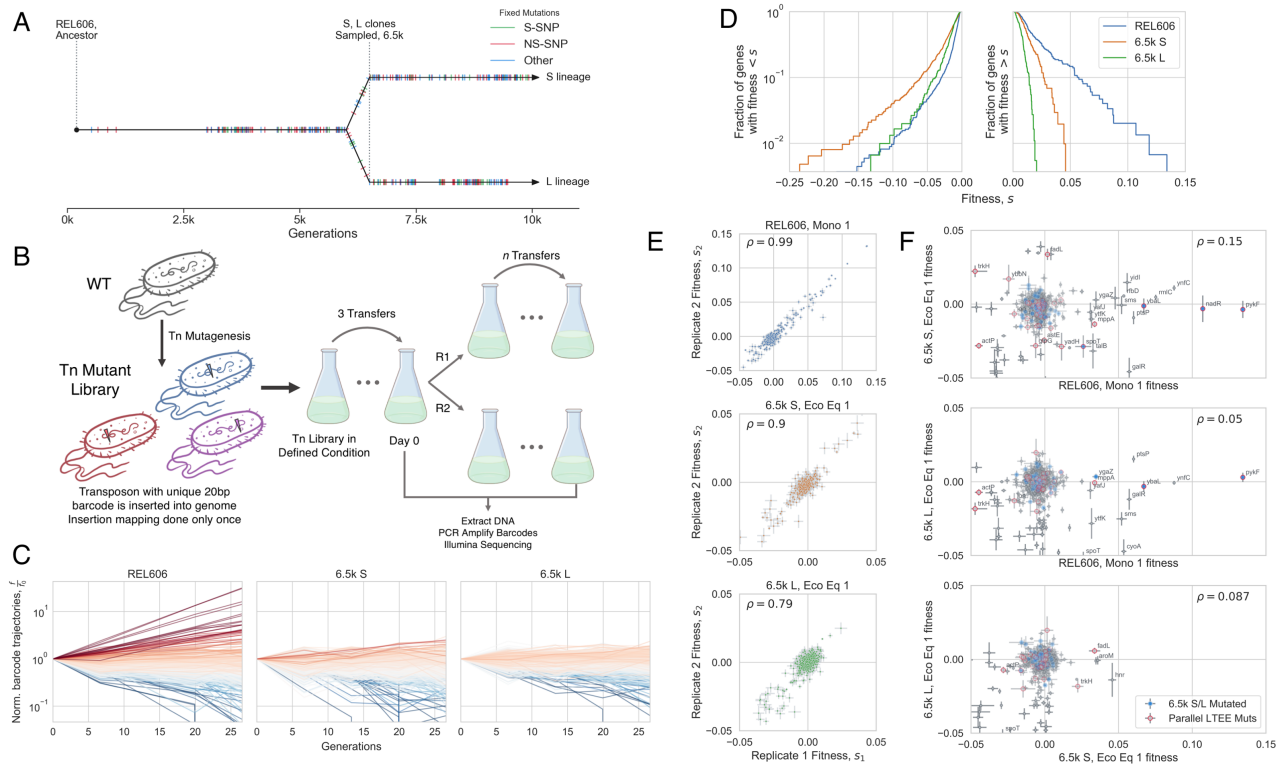


Fig. 1. Measuring mutational fitness effects. (A) Timeline of evolution in the *ara-2* LTEE population, showing mutation accumulation and diversification into S and L around 6k generations, then clone sampling at 6.5k generations; data from (25), jitter added to mutation fixation time for easier visualization. Hypermutator phenotype appeared around 2.5k generations (36). (B) Schematic of transposon mutagenesis process to generate barcoded libraries of REL606, 6.5k S and L, as well as experimental procedure to observe barcode dynamics. (C) Barcoded knockout mutant frequency trajectories in the evolutionary condition for each genetic background, colored by estimated fitness. All barcodes within a gene were summed together; shown are the trajectories from replicate 1 in the evolutionary condition for each genetic background (monoculture for REL606, together at ecological equilibrium frequency for S and L; representative of both replicates). (D) Overall distributions of fitness effects in the evolutionary condition for each genetic background. The majority of knockouts were neutral, so only genes that were called as significantly non-neutral were included (see supplement section S3.2). (E) Replicate-replicate correlation of estimated fitness effects. (F) Comparison of knockout fitness effects across genetic backgrounds, which are generally uncorrelated. Points with a blue interior correspond to genes that were mutated (excluding S-SNPs) in 6.5k S/L relative to REL606 (sequencing data from (35)). Points with red outlines correspond to genes that were mutated in parallel in nonmutator LTEE populations (data from (25)). In panels E-F, knockouts with high measurement noise ($\sigma_s > 0.3\%$) were excluded (except for labeled genes), and ρ is the weighted Pearson correlation coefficient. Also in panels E-F, the "cloud" of points around 0 mostly represents likely neutral knockouts.

205 furthermore, there are several genes that were mutated in paral- 227
 206 lel in multiple lines of the LTEE, but are only beneficial on 228
 207 the S background (*trkH*, *ybbN*) or both the S and L back- 229
 208 grounds (*fadL*) when knocked out, while being neutral or 230
 209 deleterious on the REL606 background, suggesting that pre- 231
 210 dictable epistasis could have shaped which mutations became 232
 211 beneficial in the LTEE. 'Coupon collection' is a null model of 233
 212 mutation accumulation/epistasis, where a beneficial DFE is 234
 213 composed of a finite number of mutations, and only changes 235
 214 due to the depletion of those mutations when they fix in a 236
 215 population. While the coupon collecting model is clearly rel- 237
 216 evant for some mutations, the lack of fitness effect correla- 238
 217 tion between genetic backgrounds seems to be largely driven by 239
 218 global epistasis.

219 **Knockout fitness effects strongly depend on ecologi-**
 220 **cal conditions.** The ecological interactions between S and 242
 221 L are mediated through the environment, most likely pri- 243
 222 marily through cross-feeding (31, 32). Therefore, it's rea- 244
 223 sonable to think that the environment will change with the 245
 224 ecosystem composition, which could be modified by both 246
 225 ecological and evolutionary processes—indeed, ecotype com- 247
 226 position does change significantly and relatively rapidly over

227 evolutionary time (~1k-10ks generations) (25, 30). Thus, we 228
 229 sought to explore how mutational fitness effects varied with 229
 230 ecosystem composition (Figure S6). Notably, we see a con- 230
 231 sistent trend where fitness effects generally have a smaller 231
 232 magnitude when S and L are in monoculture compared to 232
 233 when they are in coculture (Figures 2A, S10A). Addition- 233
 234 ally, we also see that the overall shape of the DFEs change 234
 235 as a function of frequency, with generally larger fitness ef- 235
 236 fects when the ecotype is in the minority, for both beneficial 236
 237 and deleterious knockouts (Figures 2B, S10B). Analogous to 237
 238 the case of global diminishing returns epistasis, this observa- 238
 239 tion holds on a statistical level, but does not explain all of the 239
 240 fitness effect variation between the different conditions, imply- 240
 241 ing that individual mutations are affected by the ecosys- 241
 242 tem composition in idiosyncratic ways—statistical properties 242
 243 of the DFE seem to be strongly dependent on the ecosys- 243
 244 tem composition, but the effects of individual mutations may 244
 245 depend on their underlying physiological consequences and 245
 246 how they affect ecological interactions. Thus, it appears that 246
 247 the impact of both ecotypes on the environment is different 247
 248 enough to make selection pressures strongly dependent on the 248
 249 current mixture of ecotypes.

The LTEE environment, while relatively simple, varies quite

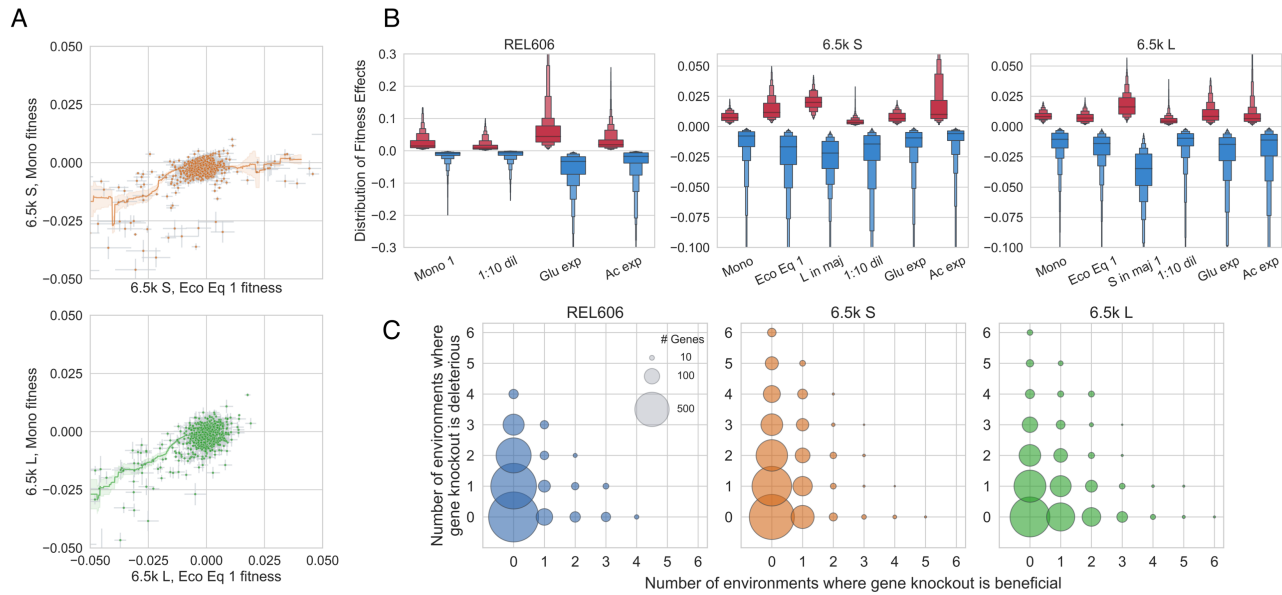


Fig. 2. Statistical properties of DFEs as well as effects of individual mutations sensitively depend on environment. (A) Knockout fitness effects tend to have a larger magnitude when S and L are at ecological equilibrium versus when they are in monoculture. Line shown is a rolling average of fitness effects \pm standard error. **(B)** Distribution of fitness effects across environments, where we only included knockouts that were called as significantly non-neutral. Please note that the DFEs of REL606 are on a different scale than S and L. **(C)** Illustration of how the sign of fitness effects changes across environments. Few mutations are unconditionally beneficial or deleterious, many are non-neutral only in one or few environments, and sign-flipping of fitness effects across environments is pervasive. REL606 only has 4 unique environments, compared to 6 for S and L. Size of circle is proportional to number of genes that fall into each class.

Experiment	Libraries	Same flasks?	Description
Mono	R, S, L	N	Library monoculture
1:10 dil	R, S, L	N	Library monoculture, 1:10 daily dilution
Glu exp	R, S, L	N	Library monoculture, kept in glucose exponential phase
Ac exp	R, S, L	N	Library monoculture, kept in acetate exponential phase
Eco Eq 1	S, L	Y	S + L libraries with wt L at ecological equilibrium
Eco Eq 2	S	N	S library with wt S + L at ecological equilibrium
Eco Eq 2	L	N	L library with wt S + L at ecological equilibrium
L in maj	S	N	S library with wt L in majority
S in maj 1	L	N	L library with wt S in majority
S in maj 2	S, L	Y	S + L libraries with wt S in majority
S in maj 3	L	N	L library with wt S in majority

Table 1. Summary of BarSeq experiments reported in this work. Dilution rate was variable in the glucose/acetate exponential phase experiments, to keep the populations in exponential phase (see supplement section S2), but unless otherwise noted, the daily dilution rate was 1:100, consistent with the LTEE condition. All experiments were performed in the LTEE media, DM25, except for the acetate exponential phase experiment. The abbreviations R, S, and L refer to REL606, and 6.5k S/L, respectively.

250 significantly over the course of a single cycle (31, 41), allowing
 251 ecotypes to carve out different temporal ecological niches
 252 during cycles of lag, exponential, and stationary phases. To
 253 explore how selection pressures vary in different niches in
 254 the growth cycle, we measured fitness in exponential growth
 255 on glucose and acetate (which appears in the LTEE environ-
 256 ment due to overflow metabolism), and at a reduced dilution
 257 rate of 1:10 such that portion of the growth cycle in station-
 258 ary phase is increased (Table 1). We found that the shape
 259 of the DFE changed substantially based on the environment
 260 (Figures 2B). For example, while S and L have a similar ben-
 261 efiticial DFE shape in monoculture, L has access to stronger
 262 beneficial knockout mutations in glucose exponential phase
 263 compared to S. As another example, the beneficial DFE in ac-
 264 etate is larger than any other DFE in both S and L, potentially
 265 pointing to a substantial, as-of-yet unrealized adaptive poten-

266 tial for adaptation on acetate. Interestingly, despite the envi-
 267 ronmental variation, REL606 always has a more pronounced
 268 beneficial DFE compared to S and L.

269 It is important to note that measurement noise varied non-
 270 negligibly across experiments, primarily because of changes
 271 in bottleneck size (and thus in the strength of genetic drift)
 272 due to differences in library frequency and and other exper-
 273 imental differences (see supplement section S2). Thus, our
 274 power to detect selected mutations close to neutrality varied
 275 across experiments. In contrast to previous work (12),
 276 it appears that there is no consistent relationship between
 277 background fitness and shape of the deleterious DFE, which
 278 instead appears to depend more on environment. Possible
 279 reasons for the discrepancy include species-dependent differ-
 280 ences, and the fact that our set of experiments used back-
 281 grounds connected by evolution, while Johnson et al. used

282 evolutionarily unrelated yeast hybrids with varying fitness in
283 the test environment, whose changing DFEs were not controlled
284 by evolution.

285 In addition to the strong dependence of the macroscopic DFE
286 on environment, it appears that the fitness effects of individual
287 mutations can also change radically by environment. Strikingly,
288 in the set of considered environments, conditional non-neutrality
289 and sign-flipping appear to be very common across all three genetic
290 backgrounds (Figure 2C). The majority of knockouts are non-neutral
291 in at least one measured environment; just about ~20% of knockouts
292 are called as neutral across all environments. Very few mutations are
293 unconditionally beneficial or deleterious across all environments,
294 and many more mutations flip signs across environments, suggesting
295 the presence of widespread trade-offs between adapting to different
296 components of an environment. Both conditional neutrality ($x > 0, y = 0$)
297 and antagonistic pleiotropy ($x > 0, y > 0$) appear prevalent in our
298 system. The ubiquitous presence of sign-flipping also suggests that
299 subtle changes to environmental conditions—e.g. by changes to ecotype
300 frequency via adaptation—could meaningfully affect evolutionary
301 outcomes by changing which mutations are likely to establish. The
302 ubiquity of sign-flipping still holds if we reduce the p-value cutoff
303 from 0.05 to 10^{-3} or 10^{-5} to determine non-neutrality (Figure
304 S11), or only consider genes with $|s| > 1\%$ or $|s| > 2\%$ as non-neutral
305 (Figure S12), although more genes are called as neutral, as would be
306 expected. However, it is important to note that we only considered
307 genes to be non-neutral if their fitness was significantly different
308 from 0; thus, it is likely that some knockouts were incorrectly
309 called as neutral, especially if their fitness effect is small. Additionally,
310 we have only measured a relatively small set of closely related
311 environments (representing niches within the LTEE environment), so
312 we might expect that if we measure fitness in a sufficiently large
313 number of environments, many more genes would be non-neutral in
314 at least one.

315 By computing the correlation of mutational fitness effects across
316 environments (weighted by measurement error), we can obtain a
317 measure of the functional similarity of environments, which we can
318 also use to cluster said environments (Figure 3). As a first
319 observation and check, it is reassuring to see the clustering of
320 quasi-replicate experiments, i.e. experiments with relatively
321 minor differences in the experimental set-up and performed on
322 different days—Eco Eq 1/2, S in maj 1/2/3 (L), and Mono 1/2
323 (REL606) (see supplement section S2). However, the correlations
324 between the quasi-replicates are lower than we see for replicate
325 experiments that we did at the same time—this could indicate
326 either that some fitness measurements are sensitive to the small
327 experimental differences (size of flasks, whether libraries are
328 cocultured or not, etc.), or simply performing the experiments
329 on different days with different environmental fluctuations leads
330 to deviations in measured fitness, as is perhaps the case in other
331 systems (14). The latter hypothesis is further supported by the
332 fact that two experiments were in fact performed at the same
333 time (S in maj 2 and 3), and had among the highest correlation of
334

all quasi-replicates.

339 Otherwise, there are still some interesting patterns that we
340 can pick out by looking at correlations across environments. For
341 example, it looks like the environments related to the putative
342 ecotype niches—glucose and acetate exponential growth in L and
343 S respectively—cluster with conditions where the ecotype is in
344 the minority. On the other hand, the monoculture experiment in
345 S clusters with glucose exponential phase. Also, in REL606 and
346 L, the acetate experiment is the outgroup compared to all the
347 other environments, and almost completely uncorrelated with
348 fitness in glucose exponential phase, but most correlated with
349 the 1:10 dilution condition. In S, this is not the case, and
350 acetate fitness is *least* correlated with 1:10 dilution fitness.
351 This may indicate that stationary phase in REL606 and L may
352 have much more acetate with which to grow on compared to
353 S, and adaptation to acetate may involve tradeoffs with
354 adaptation to glucose, at least in REL606 and L.

Correlations between genes across environments. To explore
357 the nature of the strong background dependence that we observed,
358 we sought to understand which genes are correlated with each
359 other across environments, with the intuition that genes that
360 perform the same function should change their fitness effects
361 across environments in similar ways. For example, the *sufABCDE*
362 operon encodes proteins that help to assemble iron-sulfur clusters
363 (43), and they all have correlated knockout fitness effects
364 across environments in all three genetic backgrounds (Figure
365 S14A)—as they should, if the knockouts all have very similar
366 metabolic/physiological consequences. However, other gene sets
367 are only correlated in a subset of backgrounds. Most genes in
368 the *fecABCDE* operon are correlated with each other in all
369 backgrounds except for *fecA*, which is well correlated with
370 the others in REL606, less correlated in S, and uncorrelated
371 with the others in L (Figure 4A). Similarly, the genes in the
372 *proVWX* operon are almost perfectly correlated, except one
373 condition where *proV* has a ~7% higher fitness than the other
374 two knockouts (Figure S14B). We can also look at the fitness
375 effects of a subset of knockouts that are beneficial at least
376 once for every genetic background, across environments (Figure
377 S15). We see that subsets of genes that are often beneficial
378 on a background are positively correlated with each other, e.g.
379 *pykF/cyoA* in REL606 and *ptsP/mrcA/gppA* in 6.5k L, perhaps
380 suggesting that the knockouts have common function effects.
381 These correlations often break when the mutations appear on
382 different genetic backgrounds, e.g. *pykF/cyoA* are no longer
383 correlated on (at least) the 6.5k L background, and *ptsP/mrcA/gppA*
384 are no longer correlated on the 6.5k S background, while
385 *ptsP/mrcA* actually appear *negatively* correlated on the
386 REL606 background. Together, these examples suggest that
387 correlations between knockout fitness effects may change in
388 idiosyncratic ways across genetic backgrounds.

389 We systematically quantified the pairwise correlation of
390 knockout fitness across environments—termed "cofitness",
391 previously defined in (29)—where we used the weighted
392 Pearson's correlation coefficient to account for differences in
393

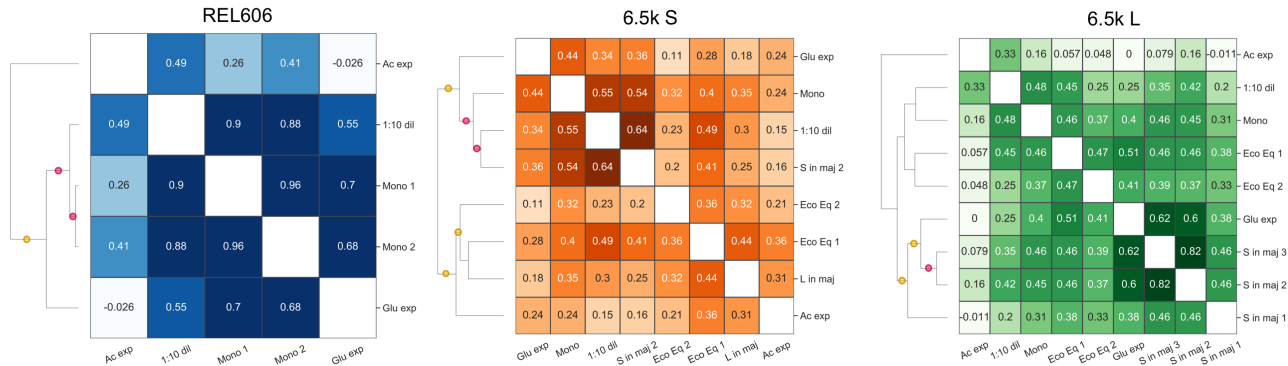


Fig. 3. Correlations of fitness effects between environments. Clustering environments by using fitness effect correlation as a measure of similarity reveals which environments are the most functionally alike. For example, environments related to the putative ecotype niches—exponential acetate growth and glucose growth, for S and L respectively—cluster with conditions where the ecotype is in the minority. The red and yellow dots indicate that the branch has $\geq 90\%$ or $\geq 70\%$ support respectively, computed via bootstrapping.

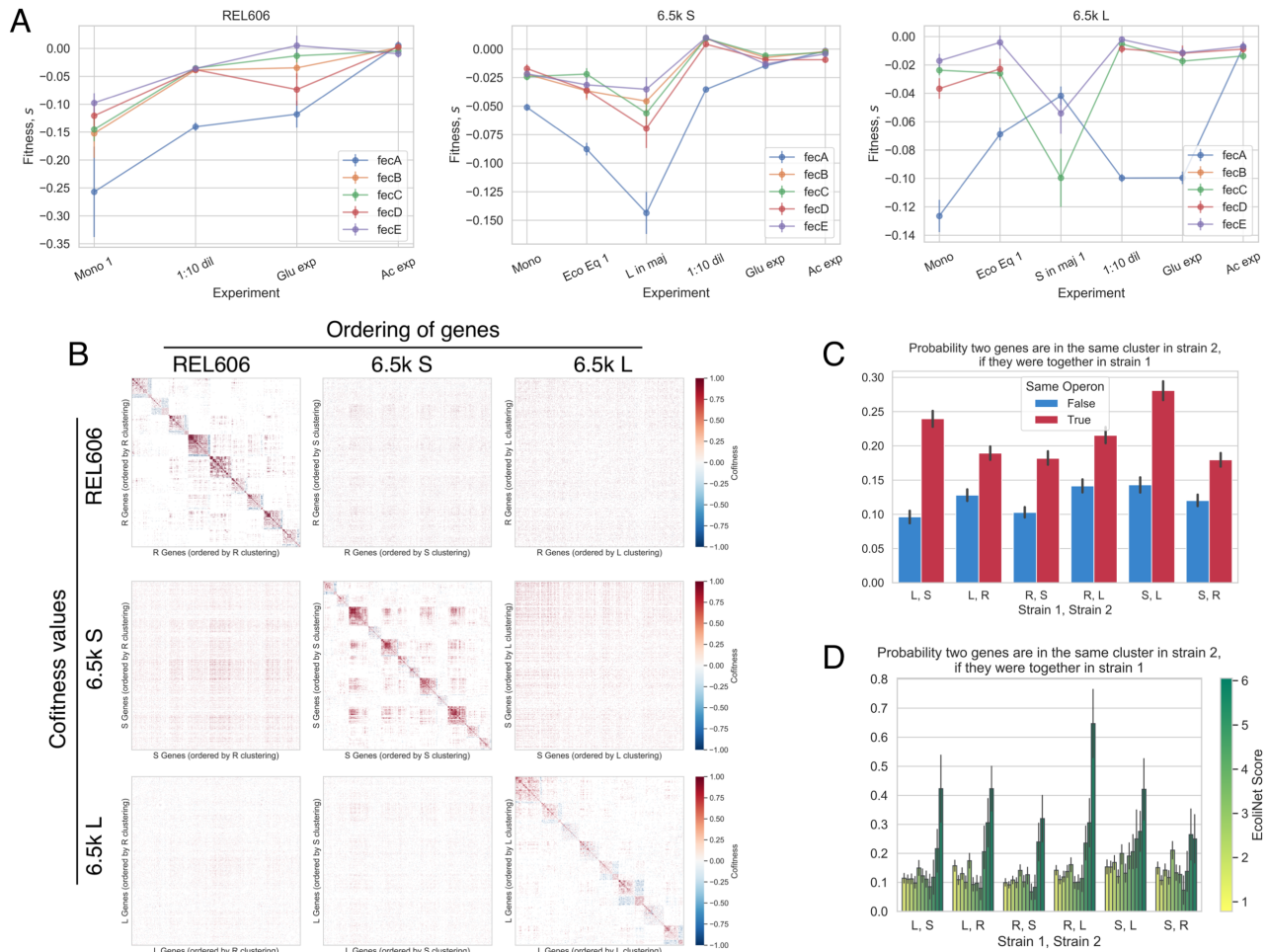


Fig. 4. Correlations between genes across environments. We observed that many pairs of genes have correlated fitness effects across environments, for example (A) most genes of the *fecABCDE* operon. However, *fecA* is correlated with the other genes to varying degrees, depending on the genetic background. (B) We computed the pairwise correlation of fitness effects (cofitness) for all pairs of genes, and then clustered genes with a community detection algorithm (42). We then rearranged the cofitness matrices by reordering genes based on “optimal” clustering of other genetic backgrounds. For each column, we ordered the genes based on the clustering of a given genetic background. For each row, we used the cofitness matrix for a given background. It is apparent that replotting the cofitness matrix using another strain’s clustering does not produce noticeable structure. (C, D) Cluster reassortment is not entirely random—pairs of genes (C) in the same operon and (D) that strongly interact with each other (high EcolNet score), tend to stay in the same clusters across genetic backgrounds. In panels C and D, the abbreviations R, S, and L refer to REL606, and 6.5k S/L, respectively.

396 measurement error across environments. We computed the
397 cofitness of all pairs of genes (excluding those called as neu-
398 tral across all environments) across the REL606, S and L li-
399 braries, as well as a null cofitness distribution for each pair
400 to determine if the two genes are significantly correlated;
401 the set of all significant gene-gene correlations determine
402 the edges in the cofitness networks (see supplement section
403 S4.2). We explored the structure of the resulting cofitness
404 networks via clustering (42) (see supplement section S4.2),
405 where we found sets of communities for all three libraries
406 with *modularity* > 0, indicating that there are more edges
407 within each community than between communities (Figure
408 4B) (44). We performed a number of controls to ensure that
409 our results weren't driven by measurement noise or technical
410 effects of clustering; see supplement section S4.2 for more
411 information.

412 The presence of strong communities suggests that most
413 knockouts are significantly correlated with others, potentially
414 pointing to similar functional effects driving changes in fit-
415 ness. We then wanted to compare how these clusters dif-
416 fer between the different genetic backgrounds, with the idea
417 that how and if clusters change should reveal information
418 on how the effective functions of genes differ across ge-
419 netic backgrounds. Surprisingly, we find that gene clusters
420 are not well preserved across genetic backgrounds, and in
421 fact, genes are typically seemingly randomly reassorted be-
422 tween genetic backgrounds (Figures 4B, S19). There are
423 a couple clusters that show non-random sampling across
424 genetic background, however, the deviation from random
425 sampling is mostly small, with one noticeable exception—
426 clusters 5, 3, and 1 in REL606, S, and L, respectively, all
427 seem to share a larger than random number of genes with
428 each other ($p < 10^{-4}$ for all clusters). From a Gene On-
429 tology enrichment analysis, genes that are associated with
430 biofilm formation (GO:0043708), adhesion (GO:0022610),
431 and pilus organization (GO:0043711) are over-represented in
432 these clusters, along with genes involved in organonitrogen
433 compound biosynthesis (GO:1901566), although to a weaker
434 extent (Figure S20). This suggests that there is at least one
435 (large) functionally related group of genes that stay corre-
436 lated across genetic backgrounds, implying that their fitness-
437 determining effects are mostly the same, regardless of genetic
438 background.

439 We wanted to know why other functional groups of genes
440 do not stay correlated with each other, and if there was any
441 structure hiding in the seeming randomness of cluster re-
442 assortment. A simple first test could ask if genes in the
443 same operon are more likely to stay correlated with each
444 other across backgrounds, which is the case for several of
445 our aforementioned examples. This indeed appears to be the
446 case across all genetic backgrounds (Figure 4C). However,
447 genes often share functions with other genes outside their
448 operons, so we turned to investigating the relationship be-
449 tween the cofitness and genetic networks. We used EcoliNet
450 as a representation of the *E. coli* genetic network, as it at-
451 tempts to capture all interactions between genes by integrat-
452 ing various data-types, regardless of the mechanism (tran-

453 scriptional, protein-protein, etc), and assigns a score to each
454 interaction that effectively represents the strength of the inter-
455 action (45). We then computed the probability that two genes
456 are in the same community in one genetic background, given
457 that they're together in another background, as a function of
458 EcoliNet score (Figures 4D). We see that gene pairs that are
459 predicted to strongly interact (high EcoliNet score) are much
460 more likely to be correlated across genetic backgrounds. We
461 can also see these same patterns without referencing any clus-
462 ter labels—if we look at the correlation between all cofitness
463 pairs across genetic backgrounds, pairs that are in the same
464 operon (Figure S21A) and those with the highest EcoliNet
465 score (Figure S21B) give the highest correlation. It also ap-
466 pears that the shortest distance between two nodes in the Eco-
467 liNet network (Figure S23) also predicts if the two genes will
468 stay correlated across genetic background, albeit the effect is
469 weaker. In all, these analyses suggest that evolution drastical-
470 ly changes which functional effects of genes are important
471 for determining fitness, such that the cofitness of genes pairs
472 is not preserved across genetic backgrounds, for all but the
473 most strongly interacting genes.

Fitness effects are correlated with evolutionary out-
474 **comes.** We sought to explore if the knockout fitness ef-
475 fects that we measured were correlated with evolutionary
476 outcomes in the LTEE, i.e. establishment of mutations and
477 changes in gene expression. So, we first investigated if genes
478 with non-neutral knockout fitness were more or less likely to
479 be mutated and rise to a sufficiently high frequency in the
480 population. Using the clonal sequencing data from Tenaillon
481 et al. (2016) (46) and Plucain et al. (2014) (35), we identi-
482 fied genes that mutated between selected LTEE time-points,
483 and ran a logistic model with fitness effect as the predictor
484 and mutated status as the response variable (see supplement
485 section S4.3), separately for beneficial (Figures 5A) and dele-
486 terious genes (S24A). We used three sequenced clones (one
487 available for each time point) for both S and L, while we used
488 all clones from all non-mutator populations (at a given time
489 point) for REL606. We used the appearance of a mutation
490 (excluding synonymous SNPs) within a gene as a proxy for
491 establishment.

492 Fitness of beneficial knockouts in the 1:10 dilution condi-
493 tion and monoculture (LTEE condition) in the REL606 back-
494 ground is strongly correlated with which mutations estab-
495 lish from 0-5k generations, while fitness in acetate exponen-
496 tial phase is only correlated with establishment later in the
497 evolution (difference in slopes between 0-5k and 5-20k is
498 significant at $p < 0.05$ for 1:10 dilution and acetate condi-
499 tions, not for monoculture or glucose conditions). This is
500 potentially a signal that the targets of selection are shifting
501 over time—REL606 may initially adapt via lag phase shorten-
502 ing/stationary phase survival, while only later adapting via in-
503 creased acetate growth rate. This could happen, for example,
504 by either clonal interference favoring the highest-effect mu-
505 tations, or due to global epistatic effects (25). The former hy-
506 pothesis is supported by the observation that three mutations
507 appear in genes with beneficial acetate knockouts at 2k gen-
508 erations, but they then disappeared by 5k generations, poten-



Fig. 5. Fitness effects of beneficial genes are correlated with evolutionary outcomes. Slopes from logistic and linear models respectively to explore if beneficial knockout fitness effects are correlated with (A) establishment of a mutation in a gene, and (B) changes in gene expression over evolutionary time, relative to neutral knockouts. Fitness effects were normalized by the median beneficial fitness effect, so that coefficients can be interpreted as the average difference in (A) log-odds establishment and (B) log-fold change in expression, between neutral knockouts and the 'typical' beneficial knockout. (A) REL606 beneficial knockout fitness is positively correlated with gene establishment probability for most environments, but in different time intervals, potentially pointing to shifting targets of selection. (B) Beneficial knockout fitness in REL606 is positively correlated with increasing gene expression over time for all environments except acetate exponential phase. In S and L, fitness in several environments—including the ecological equilibrium and acetate and glucose growth, respectively—is correlated with decreasing gene expression. Fitness in some environments is also correlated with an increased gene expression between the ancestral state and the ecotypes between 6.5k generations. Asterisks denote coefficients that are significantly different than 0 (FDR correction; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

510 tially indicating that they were out-competed by other beneficial
 511 mutations (Figure S26). There is only one S/L condition
 512 that shows a significant difference in mutation establishment
 513 probability between beneficial and neutral mutations—genes
 514 with beneficial fitness in acetate are less likely to mutate compared
 515 to neutrals in S. However, changes in gene expression
 516 suggest adaptation to acetate may be occurring through indirect
 517 routes in S, as detailed below. However, we expect our
 518 power to detect correlations between mutational fitness and
 519 mutation establishment to be lower for S and L. They have a
 520 $\sim 100x$ higher mutation rate than REL606, implying that the
 521 ratio of neutral hitchhiking to beneficial driver mutations is
 522 higher as well.

523 We also investigated if fitness effects are correlated with
 524 changes in gene expression, using microarray data from Le
 525 Gac et al. 2012 (34), which measured gene expression in
 526 REL606, and S/L at 6.5k, 17k, and 40k generations. These
 527 measurements serve as a distinct readout of evolutionary
 528 change compared to genomic mutational dynamics, because
 529 even if a gene is not directly mutated, gene expression can
 530 still change through indirect genetic interactions. Thus, gene
 531 expression measurements allow us to probe the effects of the
 532 cumulative mutations fixed by evolution, where we fit linear
 533 models to examine the relationship between knockout fitness
 534 effect of a gene and log-fold change in gene expression between
 535 two time points (see supplement section S4.4), separately for
 536 beneficial (Figure 5B) and deleterious genes (Fig-

ure S24B).

537 In REL606, genes with beneficial knockout fitness effects
 538 tend to increase in expression (relative to neutral genes) over
 539 evolutionary time; this is perhaps surprising, because we
 540 would expect selection to decrease gene expression if knock-
 541 ing out that gene is beneficial. One possibility to explain
 542 this pattern could be that most genes that are beneficial when
 543 knocked out in REL606 also have access to potentially even
 544 more highly beneficial change-of-function mutations, which
 545 may then add selection pressure to increase expression. This
 546 is the case with mutations in *pykF*, for example, where knock-
 547 ing out the gene provides a large benefit in the LTEE environ-
 548 ment, but even larger effects are available by various nonsyn-
 549 onymous mutations (47).
 550

551 In contrast, in S and L, there are a couple of environments
 552 where gene expression significantly decreases over evolu-
 553 tionary time for genes with beneficial knockout fitness ef-
 554 fects (compared to neutrals). These conditions include envi-
 555 ronments related to the putative ecotype niches—acetate and
 556 glucose exponential growth in S and L respectively. This
 557 could lead us to hypothesize again that adaptive forces in
 558 the LTEE are in the specializing direction for both ecotypes.
 559 On the other hand, while fitness in the ecological equilib-
 560 rium is associated with decreased gene expression, this is not
 561 the case for fitness in monoculture and non-specializing en-
 562 vironments, indicating again that the latter environments are
 563 less relevant for evolution in the LTEE environment. Despite

564 the fact that acetate-adapting mutations are not establishing
565 on the S background (at least initially), gene expression still
566 decreases by 40k generations, perhaps indicating that adapta-
567 tion to acetate is occurring through routes other than directly
568 mutating genes with beneficial knockout effects.
569 We also saw that S and L beneficial knockout fitness in glu-
570 cose exponential phase is *positively* correlated with an in-
571 crease in gene expression from 0-6.5k. On average, those
572 same genes decrease in relative gene expression when evolu-
573 ing on the L background, whereas they do not change on the
574 S background. This set of data could indicate that from 0-
575 6.5k many genes increased in gene expression via adaptive
576 evolution that were actively unhelpful for glucose growth, ei-
577 ther because of transcriptomic misallocation or other types
578 of antagonistic pleiotropy, such that knocking them out con-
579 ferred a benefit. Upon diversification of S and L, the direc-
580 tion of gene expression change appears to switch for L, per-
581 haps suggesting that L is evolving towards a more glucose
582 growth-optimized transcriptome, while S is not. This set of
583 observations provides a possible example of how diversifica-
584 tion changes the selection pressures acting on organisms.
585 Interestingly, deleterious knockout fitness effects across all
586 environments in S/L tend to be associated with an increase
587 in gene expression between 0 and 6.5k generations (Figure
588 S24B). This observation may provide a partial explanation
589 for why some knockouts become deleterious in S/L when
590 they were neutral in REL606–6.5k generations of evolution
591 caused the genes to suddenly become important, so they be-
592 came more costly to knock out. Another, unrelated obser-
593 vation could help us to understand why some genes have
594 deleterious knockout fitness effects—it appears that deleteri-
595 ous genes are more highly connected in the *E. coli* gene in-
596 teraction network (EcoliNet) compared to neutrals (on aver-
597 age), indicating that some genes may be deleterious because
598 when they're knockout out, they also affect the functioning
599 of many other genes (Figure S27).

600 Discussion

601 To assess how mutational fitness effects depend on genetic
602 background and ecological conditions, we measured the
603 genome-wide knockout fitness effects of a recently diversi-
604 fied ecosystem, S and L, and their ancestor, REL606. Despite
605 the fact that the fitness effects of individual mutations appear
606 to be highly dependent on both genetic background and envi-
607 ronment (strong $(G \times) G \times E$ effects), we saw consistent sta-
608 tistical patterns of variation across both axes, namely global
609 diminishing returns epistasis and a negative frequency-fitness
610 correlation (in S and L). In contrast, previous studies that ob-
611 served diminishing returns epistasis saw both the mean of the
612 DFE as well as the fitness effects of individual mutations de-
613 crease as a function of background fitness (39, 40); this dis-
614 crepancy may indicate that uniform negative epistasis of in-
615 dividual mutations may only be relevant for the first handful
616 of mutational steps, before yielding to more complex and id-
617 iosyncratic forms of epistasis.
618 Even though S and L only diverged ~500 generations ago, the
619 mixing ratio of the two ecotypes strongly affects the DFEs,

620 suggesting that strong eco-evolutionary coupling is possible
621 even in closely related strains. This would imply that selec-
622 tive pressures depend strongly on the community mixture,
623 which changes significantly and relatively rapidly due to evo-
624 lution (25, 30). The sensitivity of knockout fitness effects to
625 relatively minor variations on the LTEE environment, such as
626 changing niche availability or ecosystem composition, may
627 be evolutionarily significant—we know that the growth traits
628 of S and L also change quite drastically during their coevolu-
629 tion (32, 34), which along with changes to ecosystem compo-
630 sition, will change the environment, and thus change which
631 mutations are favored by selection. One specific hypothe-
632 sis that emerges from our data is that selective forces may
633 be more similar to environments related to the putative eco-
634 type niches when the ecotype is rare, for both S and L. This
635 is supported by both clustering environments by fitness ef-
636 fect correlations, and which environments were correlated
637 with changes in gene expression. It would follow that se-
638 lection could favor different degrees of specialization within
639 the current niche as the ecotype frequencies and growth traits
640 change due to evolution. Regardless of the specific imple-
641 mentation, the process where (i) mutations change growth
642 traits and ecosystem composition, which (ii) change ecolog-
643 ical conditions, which in turn (iii) change the mutational fit-
644 ness effects of both ecotypes, could represent an important
645 and pervasive type of eco-evolutionary feedback.

646 We aimed to better understand the background and environ-
647 ment dependence of mutational fitness effects by systemati-
648 cally studying fitness correlations across environments. Our
649 intuition was that knocking out genes with similar functions
650 should have similar effects across environments. We saw that,
651 by and large, different sets of genes were correlated with
652 each other across genetic backgrounds; only strongly inter-
653 acting pairs of genes were likely to be correlated across
654 all backgrounds. These widespread changes could be caused
655 by a number of different evolutionary phenomena—for exam-
656 ple, evolution could have induced widespread changes in the
657 functional effects of genes or which functional effects matter
658 for fitness. Additionally, inasmuch as fitness in an environ-
659 ment is a reflection of phenotype—e.g. fitness in exponential
660 phase is likely a simple function of exponential growth rate—the
661 extensive changes in fitness across environments could be
662 interpreted as support for ubiquitous pleiotropic effects of
663 knockout mutations.

664 We investigated if our measured knockout fitness effects were
665 correlated with evolutionary outcomes, i.e. mutation estab-
666 lishment and gene expression changes. We found significant
667 correlations across several, but not all environments, lead-
668 ing to hypotheses on how selection has acted on LTEE pop-
669 ulations. From correlations of knockout fitness effects with
670 mutation establishment, we found potential signals of shift-
671 ing selection over time in REL606. Changes in gene expres-
672 sion provide a distinct window into evolutionary change, as
673 expression can change through genetic interactions, even if
674 a gene is not directly mutated. Among other patterns, the
675 fitness correlations with gene expression changes potentially
676 reveal how the traits under selection changed from pre- to

677 post-diversification, and how they are different between S and
678 L. Overall, the connections between evolutionary changes
679 and knockout fitness effects demonstrates the utility of our
680 approach to understand how adaptation happens in the "natu-
681 ral" evolutionary context.

682 Our study is limited by the fact that we only surveyed the fit-
683 ness effects of knockout mutations, which represent a subset
684 of all mutations available to an organism. While it is possible
685 that other types of mutations could display different patterns,
686 knockout mutations appear to be prevalent and important for
687 adaptation in the LTEE (25, 48), and our measured knock-
688 out fitness effects are correlated with evolutionary outcomes.
689 Additionally, we studied a relatively simple ecosystem, con-
690 sisting of just two recently diverged ecotypes; measuring the
691 mutational effects in more complicated ecosystems and how
692 they change as a result of longer periods of evolution is likely
693 a fruitful future avenue of investigation. Overall, the meth-
694 ods and results presented here pave the way for future studies
695 investigating how mutational fitness effects depend on eco-
696 evolutionary processes, and how eco-evolutionary feedback
697 arises from changing fitness effects.

698 Methods

699 See supplementary information.

700 Data and code availability

701 All data and code used to process the data and perform the
702 analyses are available on GitHub,
703 <https://github.com/joaoscensao/S-L-REL606-BarSeq>

704 ACKNOWLEDGEMENTS

705 We thank Matti Gralka and QinQin Yu for early-phase experimental assistance.
706 We thank current and former members of the Hallatschek and Arkin labs for pro-
707 ductive discussions and feedback regarding this project, especially Morgan Price,
708 Jonas Denk, and QinQin Yu. We thank Richard Lenski for sending us the REL606,
709 6.5k S and L clones, along with experimental advice and feedback. Research re-
710 ported in this publication was supported by a National Science Foundation CA-
711 REER Award (1555330) and by the Miller Institute for Basic Research in Science,
712 University of California, Berkeley. Some elements of this project were funded by
713 ENIGMA-Ecosystems and Networks Integrated with Genes and Molecular Assem-
714 blies (<http://enigma.lbl.gov>), a Science Focus Area Program at Lawrence Berke-
715 ley National Laboratory, and is based upon work supported by the U.S. Depart-
716 ment of Energy, Office of Science, Office of Biological & Environmental Research
717 under contract number DE-AC02-05CH11231. JAA acknowledges support from
718 an NSF graduate research fellowship, a Berkeley fellowship (from UC Berkeley),
719 and Lloyd and Brodie scholarships (from UC Berkeley Dept of Bioengineering).
720 BHG acknowledges support from an award from the Alfred P. Sloan Founda-
721 tion (FG-2021-15708). This research used resources of the National Energy Re-
722 search Scientific Computing Center (NERSC), a U.S. Department of Energy Of-
723 fice of Science User Facility located at Lawrence Berkeley National Laboratory,
724 operated under Contract No. DE-AC02-05CH11231. This work used the Vin-
725 cent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by
726 NIH S10 OD018174 Instrumentation Grant. The bacteria (IGI prokaryote icon;
727 <https://innovativegenomics.org/glossary/>) and erlenmeyer flask (flask-2 icon by DB-
728 CLS; <https://bioicons.com/>) clipart from Figure 1B are used and modified under cre-
729 ative commons licenses, CC BY-NC-SA 4.0 and CC-BY 4.0 respectively. BioRxiv
730 preprint was prepared in Overleaf with the (modified) Henriques Lab LaTeX tem-
731 plate.

732 References

- 733 1. John H. Gillespie. Molecular Evolution Over the Mutational Landscape. *Evolution*, 38(5):
734 1116, 9 1984. doi: 10.2307/2408444.
- 735 2. H. Allen Orr. The Population Genetics of Adaptation: The Distribution of Factors Fixed
736 during Adaptive Evolution. *Evolution*, 52(4):935, 8 1998. doi: 10.2307/2411226.
- 737 3. H. Allen Orr. The distribution of fitness effects among beneficial mutations. *Genetics*, 163
738 (4):1519, 4 2003.

4. Benjamin H. Good, Igor M. Rouzine, Daniel J. Balick, Oskar Hallatschek, and Michael M. Desai. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of the National Academy of Sciences of the United States of America*, 109(13):4950–4955, 2012. ISSN 00278424. doi: 10.1073/pnas.1119910109. 742
5. Daniel P Rice, Benjamin H Good, Michael M Desai, and R. Korona. The evolutionarily stable distribution of fitness effects. *Genetics*, 200(1):321–9, 5 2015. ISSN 1943-2631. doi: 10.1534/genetics.114.173815. 743
6. Michael M Desai and Daniel S Fisher. Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–98, 7 2007. ISSN 0016-6731. doi: 10.1534/genetics.106.067678. 744
7. Sasha F. Levy, Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181–186, 3 2015. ISSN 0028-0836. doi: 10.1038/nature14279. 745
8. Alex N. Nguyen Ba, Ivana Cvijović, José I. Rojas Echenique, Katherine R. Lawrence, Artur Rego-Costa, Xianan Liu, Sasha F. Levy, and Michael M. Desai. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* 2019, pages 1–11, 2019. ISSN 0028-0836. doi: 10.1038/s41586-019-1749-3. 746
9. Kimberly S. Vasquez, Lisa Willis, Nate J. Cirra, Katharine M. Ng, Miguel F. Pedro, Andrés Aranda-Díaz, Manohary Rajendram, Feiqiao Brian Yu, Steven K. Higginbottom, Norma Neff, Gavin Sherlock, Karina B. Xavier, Stephen R. Quake, Justin L. Sonnenburg, Benjamin H. Good, and Kerwyn Casey Huang. Quantifying rapid bacterial evolution and transmission within the mouse intestine. *Cell Host & Microbe*, 29(9):1454–1468, 9 2021. ISSN 1931-3128. doi: 10.1016/j.chom.2021.08.003. 747
10. Jamie R. Blundell, Katja Schwartz, Danielle Francois, Daniel S. Fisher, Gavin Sherlock, and Sasha F. Levy. The dynamics of adaptive genetic diversity during the early stages of clonal evolution. *Nature Ecology & Evolution*, 3(2):293–301, 2 2019. ISSN 2397-334X. doi: 10.1038/s41559-018-0758-1. 748
11. Sandeep Venkataram, Barbara Dunn, Yuping Li, Atish Agarwala, Jessica Chang, Emily R Ebel, Kerry Geiler-Samerotte, Lucas Hérisant, Jamie R Blundell, Sasha F Levy, Daniel S Fisher, Gavin Sherlock, and Dmitri A Petrov. Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell*, 166(6):1585–1596, 9 2016. ISSN 1097-4172. doi: 10.1016/j.cell.2016.08.002. 749
12. Milo S. Johnson, Alena Martsul, Sergey Kryazhinskiy, and Michael M. Desai. Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science*, 366(6464):490–493, 10 2019. ISSN 10959203. doi: 10.1126/science.aay4199. 750
13. Celia Payen, Anna B. Sunshine, Giang T. Ong, Jamie L. Pogachar, Wei Zhao, and Maitreya J. Dunham. High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations. *PLoS Genetics*, 12(10):e1006339, 10 2016. ISSN 1553-7404. doi: 10.1371/JOURNAL.PGEN.1006339. 751
14. Grant Kinsler, Kerry Geiler-Samerotte, and Dmitri Petrov. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife*, 9:1–52, 12 2020. doi: 10.7554/ELIFE.61271. 752
15. Yuping Li, Sandeep Venkataram, Atish Agarwala, Barbara Dunn, Dmitri A. Petrov, Gavin Sherlock, and Daniel S. Fisher. Hidden Complexity of Yeast Adaptation under Simple Evolutionary Conditions. *Current Biology*, 28(4):515–525, 2 2018. ISSN 09609822. doi: 10.1016/j.cub.2018.01.009. 753
16. Yuping Li, Dmitri A. Petrov, and Gavin Sherlock. Single nucleotide mapping of trait space reveals Pareto fronts that constrain adaptation. *Nature Ecology and Evolution*, 3(11):1539–1551, 2019. ISSN 2397334X. doi: 10.1038/s41559-019-0993-0. 754
17. Shijie Zhao, Tami D. Lieberman, Mathilde Poyet, Kathryn M. Kauffman, Sean M. Gibbons, Mathieu Groussin, Ramnik J. Xavier, and Eric J. Alm. Adaptive Evolution within Gut Microbiomes of Healthy People. *Cell Host and Microbe*, 25(5):656–667, 2019. ISSN 19346069. doi: 10.1016/j.chom.2019.03.007. 755
18. Anders Folkesson, Lars Jelsbak, Lei Yang, Helle Krogh Johansen, Oana Ciofu, Niels Hoiby, and Soren Molin. Adaptation of *Pseudomonas aeruginosa* to the cystic fibrosis airway: An evolutionary perspective, 12 2012. ISSN 17401526. 756
19. Ana Sousa, Ricardo S. Ramiro, João Barroso-Batista, Daniela Güleresi, Marta Lourenço, and Isabel Gordo. Recurrent Reverse Evolution Maintains Polymorphism after Strong Bottlenecks in Commensal Gut Bacteria. *Molecular Biology and Evolution*, 34(11):2879–2892, 11 2017. ISSN 0737-4038. doi: 10.1093/MOLBEV/MSX221. 757
20. Evgenii M. Frenkel, Michael J. McDonald, J. David Van Dyken, Katya Kosheleva, Gregory I. Lang, and Michael M. Desai. Crowded growth leads to the spontaneous evolution of semistable coexistence in laboratory yeast populations. *Proceedings of the National Academy of Sciences of the United States of America*, 112(36):11306–11311, 2015. ISSN 10916490. doi: 10.1073/pnas.1506184112. 758
21. Z. D. Blount, C. Z. Borland, and R. E. Lenski. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 105(23):7899–7906, 6 2008. ISSN 0027-8424. doi: 10.1073/pnas.0803151105. 759
22. Matthew D. Herron and Michael Doebeli. Parallel Evolutionary Dynamics of Adaptive Diversification in *Escherichia coli*. *PLoS Biology*, 11(2):e1001490, 2 2013. ISSN 1545-7885. doi: 10.1371/journal.pbio.1001490. 760
23. Kinnersley M, Wenger J, Kroll E, Adams J, Sherlock G, and Rosenzweig F. Ex uno plures: clonal reinforcement drives evolution of a simple microbial community. *PLoS genetics*, 10(6), 2014. ISSN 1553-7404. doi: 10.1371/JOURNAL.PGEN.1004430. 761
24. Helling RB, Vargas CN, and Adams J. Evolution of *Escherichia coli* during growth in a constant environment. *Genetics*, 116(3):349–358, 7 1987. ISSN 0016-6731. doi: 10.1093/GENETICS/116.3.349. 762
25. Benjamin H. Good, Michael J. McDonald, Jeffrey E. Barrick, Richard E. Lenski, and Michael M. Desai. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678):45–50, 10 2017. ISSN 0028-0836. doi: 10.1038/nature24287. 763
26. Benjamin H Good, Stephen Martis, and Oskar Hallatschek. Adaptation limits ecological diversification and promotes ecological tinkering during the competition for substitutable resources. *Proceedings of the National Academy of Sciences of the United States of America*, 115(44):E10407–E10416, 10 2018. ISSN 1091-6490. doi: 10.1073/pnas.1807530115. 764
27. Massimo Amicone and Isabel Gordo. Molecular signatures of resource competition: Clonal 765

- interference favors ecological diversification and can lead to incipient speciation. *Evolution*, 2021. ISSN 1558-5646. doi: 10.1111/EVO.14315.
28. Kelly M Wetmore, Morgan N Price, Robert J Waters, Jacob S Lamson, Jennifer He, Cindi A Hoover, Matthew J Blow, James Bristow, Gareth Butland, Adam P Arkin, and Adam Deutschbauer. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio*, 6(3):00306–15, 5 2015. ISSN 2150-7511. doi: 10.1128/mBio.00306-15.
29. Morgan N. Price, Kelly M. Wetmore, R. Jordan Waters, Mark Callaghan, Jayashree Ray, Hualan Liu, Jennifer V. Kuehl, Ryan A. Melnyk, Jacob S. Lamson, Yumi Suh, Hans K. Carlson, Zuelma Esquivel, Harini Sadeeshkumar, Romy Chakraborty, Grant M. Zane, Benjamin E. Rubin, Judy D. Wall, Axel Visel, James Bristow, Matthew J. Blow, Adam P. Arkin, and Adam M. Deutschbauer. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*, page 1, 5 2018. ISSN 0028-0836. doi: 10.1038/s41586-018-0124-0.
30. Daniel E Rozen and Richard E Lenski. Long-Term Experimental Evolution in *Escherichia coli*. VIII. Dynamics of a Balanced Polymorphism. *The American Naturalist*, 155(1):24–35, 1 2000. ISSN 1537-5323. doi: 10.1086/303299.
31. Daniel E. Rozen, Nadège Philippe, J. Arjan de Visser, Richard E. Lenski, and Dominique Schneider. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecology Letters*, 12(1):34–44, 1 2009. ISSN 1461023X. doi: 10.1111/j.1461-0248.2008.01257.x.
32. Tobias Großkopf, Jessika Consuegra, Joël Gaffé, John C. Willison, Richard E. Lenski, Orkun S. Soyer, and Dominique Schneider. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. *BMC Evolutionary Biology*, 16(1):163, 12 2016. ISSN 1471-2148. doi: 10.1186/s12862-016-0733-x.
33. Daniel E. Rozen, Dominique Schneider, and Richard E. Lenski. Long-Term Experimental Evolution in *Escherichia coli*. XIII. Phylogenetic History of a Balanced Polymorphism. *Journal of Molecular Evolution*, 61(2):171–180, 8 2005. ISSN 0022-2844. doi: 10.1007/s00239-004-0322-2.
34. Mickaël Le Gac, Jessica Plucain, Thomas Hindré, Richard E. Lenski, and Dominique Schneider. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 96(18):10242–7, 8 2012. ISSN 0027-8424. doi: 10.1073/pnas.96.18.10242.
35. Jessica Plucain, Thomas Hindré, Mickaël Le Gac, Olivier Tenaille, Stéphane Cruveiller, Claudine Médigue, Nicholas Leiby, William R. Harcombe, Christopher J. Marx, Richard E. Lenski, and Dominique Schneider. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science*, 343(6177):1366–1369, 2014. ISSN 10959203. doi: 10.1126/science.1248688.
36. Paul D. Sniegowski, Philip J. Gerrish, and Richard E. Lenski. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature*, 387(6634):703–705, 1997. ISSN 00280836. doi: 10.1038/42701.
37. Benjamin H. Good and Michael M. Desai. The Impact of Macroscopic Epistasis on Long-Term Evolutionary Dynamics. *Genetics*, 199(1):177–190, 1 2015. ISSN 0016-6731. doi: 10.1534/genetics.114.172460.
38. Michael J. Wiser, Noah Ribeck, and Richard E. Lenski. Long-term dynamics of adaptation in asexual populations. *Science*, 342(6164):1364–1367, 2013. ISSN 10959203. doi: 10.1126/science.1243357.
39. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science*, 332(6034):1193–1196, 6 2011. ISSN 0036-8075. doi: 10.1126/science.1203801.
40. Sergey Kryazhimskiy, Daniel P Rice, Elizabeth R Jerison, and Michael M Desai. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science (New York, N.Y.)*, 344(6191):1519–1522, 6 2014. ISSN 1095-9203. doi: 10.1126/science.1250939.
41. Farida Vasi, Michael Travisano, and Richard E. Lenski. Long-Term Experimental Evolution in *Escherichia coli*. II. Changes in Life-History Traits During Adaptation to a Seasonal Environment. <https://doi.org/10.1086/285685>, 144(3):432–456, 10 1994. doi: 10.1086/285685.
42. Ferran Parés, Dario Garcia-Gasulla, Armand Vilalta, Jonatan Moreno, Eduard Ayguadé, Jesús Labarta, Ulises Cortés, and Toyotaro Suzumura. Fluid Communities: A Competitive, Scalable and Diverse Community Detection Algorithm. *Studies in Computational Intelligence*, 689:229–240, 3 2017.
43. Yasuhiro Takahashi and Umechiyo Tokumoto. A Third Bacterial System for the Assembly of Iron-Sulfur Clusters with Homologs in Archaea and Plastids. *Journal of Biological Chemistry*, 277(32):28380–28383, 8 2002. ISSN 0021-9258. doi: 10.1074/JBC.C200365200.
44. Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Görke, Martin Hoefler, Zoran Nikoloski, and Dorothea Wagner. On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20(2):172–188, 2 2008. ISSN 10414347. doi: 10.1109/TKDE.2007.190689.
45. Kim H, Shim JE, Shin J, and Lee I. EcoliNet: a database of cofunctional gene network for *Escherichia coli*. *Database : the journal of biological databases and curation*, 2015, 2015. ISSN 1758-0463. doi: 10.1093/DATABASE/BAV001.
46. Olivier Tenaille, Jeffrey E. Barrick, Noah Ribeck, Daniel E. Deatherage, Jeffrey L. Blanchard, Aurko Dasgupta, Gabriel C. Wu, Sébastien Wielgoss, Stéphane Cruveiller, Claudine Médigue, Dominique Schneider, and Richard E. Lenski. Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature*, 536(7615):165–170, 8 2016. ISSN 14764687. doi: 10.1038/nature18959.
47. Peng F, Widmann S, Wünsche A, Duan K, Donovan KA, Dobson RCJ, Lenski RE, and Cooper TF. Effects of Beneficial Mutations in *pykF* Gene Vary over Time and across Replicate Populations in a Long-Term Experiment with Bacteria. *Molecular biology and evolution*, 35(1):202–210, 1 2018. ISSN 1537-1719. doi: 10.1093/MOLBEV/MSX279.
48. Rohan Maddamsetti, Philip J. Hatcher, Anna G. Green, Barry L. Williams, Debora S. Marks, and Richard E. Lenski. Core Genes Evolve Rapidly in the Long-Term Evolution Experiment with *Escherichia coli*. *Genome Biology and Evolution*, 9(4):1072, 4 2017. doi: 10.1093/GBE/ EVX064.
49. Antti Haapala, Esa Määttä, Mikko Ohtamaa, and David Necas. python-Levenshtein.
50. Murray Aitkin, Richard J. Boys, and Tom Chadwick. Bayesian point null hypothesis testing via the posterior likelihood ratio. *Statistics and Computing*, 15(3):217–230, 2005. ISSN 09603174. doi: 10.1007/s11222-005-1310-0.
51. Isabelle Smith and André Ferrari. Equivalence between the posterior distribution of the likelihood ratio and a p-value in an invariant frame. *Bayesian Analysis*, 9(4):939–962, 2014. ISSN 19316690. doi: 10.1214/14-BA877.
52. Peter J. Rousseeuw and Annick M. Leroy. *Robust Regression and Outlier Detection*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA, 10 1987. ISBN 9780471852339. doi: 10.1002/0471725382.
53. F. Lemoine, J.-B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, T. De Oliveira, and O. Gascuel. Renewing Felsenstein’s phylogenetic bootstrap in the era of big data. *Nature* 2018 556:7702, 556(7702):452–456, 4 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0043-0.
54. Aric A Hagberg, Daniel A Schult, and Pieter J Swart. Exploring Network Structure, Dynamics, and Function using NetworkX. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11–15, Pasadena, CA USA, 2008.
55. D. V. Klopfenstein, Liangsheng Zhang, Brent S. Pedersen, Fidel Ramirez, Alex Warwick Vesztrocy, Aurélien Naldi, Christopher J. Mungall, Jeffrey M. Yunes, Olga Botvinnik, Mark Weigel, Will Dampier, Christophe Dessimoz, Patrick Flick, and Haibao Tang. GO-TOOLS: A Python library for Gene Ontology analyses. *Scientific Reports*, 8(1):10872, 12 2018. ISSN 2045-2322. doi: 10.1038/s41598-018-28948-z.
56. Benjamin H. Good, Yves Alexandre De Montjoye, and Aaron Clauset. Performance of modularity maximization in practical contexts. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 81(4):046106, 4 2010. ISSN 15393755. doi: 10.1103/PHYSREVE.81.046106/FIGURES/14/MEDIUM.
57. Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
58. NCBI GEO. Evolution of gene expression during long term coexistence in a bacterial evolution experiment, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE30639>.
59. Davis S and Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics (Oxford, England)*, 23(14):1846–1847, 7 2007. ISSN 1367-4811. doi: 10.1093/BIOINFORMATICS/BTM254.
60. G. K. Smyth. limma: Linear Models for Microarray Data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, pages 397–420, 12 2005. doi: 10.1007/0-387-29362-0_{_}23.
61. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology*, 3(1), 2004. ISSN 1544-6115. doi: 10.2202/1544-6115.1027.

Supplement

Contents

951	S1 Barcoded transposon library construction	12
952	S2 BarSeq experiments	12
953	S2.1 Set-up of experiments	12
954	S2.2 Conditions for each experiment	13
955	S2.2.1 Monoculture	13
956	S2.2.2 Coculture experiments	13
957	S2.2.3 1:10 dilution	13
958	S2.2.4 Acetate exponential phase	13
959	S2.2.5 Glucose exponential phase	14
960	S2.3 DNA extraction, PCR, Sequencing	14
961	S3 Fitness inference pipeline	14
962	S3.1 Read counting and error correction	14
963	S3.2 Probabilistic model of read count trajectories and fitness inference	15
964	S3.3 Estimation of error parameters	15
965	S3.4 Estimation of mean fitness dynamics	16
966	S3.5 Identification of putative outlier barcodes	17
967	S3.5.1 Simulations	17
968	S3.6 Consequences of potential barcode frequency biases	17
969		
970		
971	S4 Analysis	18
972	S4.1 Correlation of fitness effects across environments	18
973	S4.2 Network of gene-by-gene correlations	19
974	S4.3 Genome evolution	20
975	S4.4 Changes in gene expression	20
976		
977		
978		
979		
980		
981		
982		
983		
984		
985		
986		
987		
988		
989		
990		
991		
992		
993		
994		
995		
996		
997		
998		

S18	Cofitness network comparison, across biological replicates	31	999
S19	Cofitness cluster reassignment	32	1000
S20	Biofilm-associated genes in the same clusters	33	1001
S21	Correlations between cofitness across strains	33	1002
S22	Variance in fitness effect across environment and EcoliNet score.	34	1003
S23	Shortest (graph) distance between genes in EcoliNet predicts if genes stay correlated across genetic background	34	1004
S24	Relationship between deleterious mutations and evolutionary outcomes	35	1005
S25	Fitness-gene expression change relationship	36	1006
S26	Establishment of mutation by knockout fitness	37	1007
S27	Fitness effects predict EcoliNet node degree	37	1008

S1 Barcoded transposon library construction 1014

To construct the barcoded transposon libraries, we isolated subclones of REL606, REL11555 (6.5k S), and REL11556 (6.5k L), all gifts of Richard Lenski (Michigan State University). Transposon mutagenesis was performed as previously described (28, 29) by mating each LTEE clone with an *E. coli* WM3064 donor (Diaminopimelic acid [DAP] auxotroph and pir⁺) containing previously described (28) randomly barcoded Tn5 plasmids with a kanamycin cassette and an R6K origin of replication. The LTEE clones were grown in DM2000 (Davis Minimal Media with 2000mg/L D-glucose), and the donor was grown in LB/Kan, all to mid-log phase. After washing the cultures, each LTEE culture was then mixed with the donor in a 1:1 ratio, then placed on 0.45 μM nitrocellulose filters (Millipore cat. no. HAWP04700) on top of a 1% agar plate with EZ-MOPS rich, defined media (Teknova cat. no. M2105) + 20mM sodium pyruvate ('EZ-py') + 0.3mM DAP. The rich media was chosen because it had a number of different carbon sources (glucose, amino acids, pyruvate) and sufficient amounts of all other required macro/micronutrients, lessening the chances of substantial negative selection in the growth media. After conjugation, the filters were picked up and placed in rich media; subsequently, the resuspended cells were plated on EZ-py agar plates supplemented with 50 μg/mL kanamycin. After approximately 24hrs of growth at 37C, colonies were scraped up and grown in EZ-py liquid media with 50 μg/mL kanamycin until OD~1; we then saved the cultures in several 10% glycerol stocks. Transposon insertion mapping (Tn-Seq) libraries were prepared as previously described (28); libraries were then sequenced on the Illumina HiSeq 4000 (150PE) at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. The resulting sequencing data was used to create a table relating each barcode to a genomic insertion location, using a previously developed script (MapTnSeq.pl) (28).

S2 BarSeq experiments 1050

S2.1 Set-up of experiments. To start a BarSeq experiment, we first unfroze 1mL glycerol stock of the REL606,

1053 6.5k S and/or 6.5k L transposon libraries and transferred
1054 the entirety to 10mL EZ-py media (media used for library
1055 construction) in 50mL glass erlenmeyer flasks, which were
1056 grown for 16-24hrs at 37C, shaken at 120rpm. All cultures
1057 for all experiments were grown with the same shaker, in the
1058 same 37C warm room. In several experiments where we measured
1059 fitness effects of 6.5k S/L barcoded libraries at various
1060 ecotype frequencies, we also grew the wild type S/L with the
1061 same media, under the same conditions. The next day, we
1062 washed the cultures by pelleting via centrifugation for 3 minutes
1063 at 5000rpm, aspirating the supernatant, and resuspending
1064 in DM0 (Davis Minimal Media without a carbon source)
1065 three times. After thoroughly vortexing the cultures, we
1066 transferred them 1:1000 to the appropriate media in n flasks
1067 (see below)—depending on the experiment, we used different
1068 numbers of flasks and different sizes, either 10mL media in
1069 50mL glass flasks or 200mL media in 1L glass flasks (same
1070 ratios, scaled up). We used multiple flasks and larger flasks to
1071 increase the total population size, decreasing fluctuations due
1072 to genetic drift. We then performed two more transfers in the
1073 appropriate conditions for the experiment to help physiologically
1074 adapt the cultures to the conditions. If we were doing
1075 a coculture experiment, we would mix the cultures at the appropriate
1076 frequencies during the second transfer. If we used
1077 multiple flasks in an experiment, we would sample an equal
1078 amount of culture from each flask into a microcentrifuge or
1079 Falcon tube, thoroughly mix the cultures, and redistribute
1080 among the same number of flasks with new media—thus, the
1081 cultures distributed in multiple flasks were effectively all part
1082 of the same population. After the third transfer, we would
1083 collect cells for day 0 of the experiment, and use that culture
1084 to start two biological replicates that are independently propagated
1085 for the remainder of the experiment. All cultures were
1086 grown at 37C, shaken at 120rpm. Cells were harvested at
1087 defined time points by centrifugation at 15000rpm for 10min
1088 of ~60mL culture for all experiments except Ac Exp (10mL)
1089 and Mono 2 (30mL), pooling culture from all flasks in an
1090 experiment/replicate at equal ratios. Subsequently, the pellets
1091 were stored at -80C until the experiment was finished.

1092 **S2.2 Conditions for each experiment.**

1093 **S2.2.1 Monoculture.** For the Mono (1) experiments, we propagated
1094 the libraries alone in DM25 (Davis Minimal Media with 25mg/L D-glucose)
1095 in 5x 50mL flasks over the course of 4 days. For the REL606 Mono 2
1096 experiment, we used 3x 50mL flasks over the course of 8 days, with
1097 four biological replicates in DM25. We transferred cultures 1:100
1098 every 24hrs, and took the number of generations per transfer as
1099 $\log_2 100$.
1100

1101 **S2.2.2 Coculture experiments.** As mentioned above, we started
1102 wildtype cultures of 6.5k S and/or L clones (same clones used to
1103 make the RB-Tn libraries) at the same time and with the same
1104 procedure as the library cultures (Table 1, main text), and mixing
1105 the cultures at the appropriate frequencies at the second "adaptation"
1106 serial transfer. We measured the ecological equilibrium frequency
1107 to be approximately 15 – 20%

1108 S (Figure S7), so we ensured that the S frequency was started
1109 in that range for the "ecological equilibrium" experiments. We
1110 started the "S/L in majority" experiments such that the minority
1111 ecotype was > 10% of the total population (Figure S6).
1112

1113 We used DM25 media and propagated the cultures for 4 days,
1114 except for S in maj 2/3 where we used 6 days, transferring
1115 1:100 every 24hrs ($\log_2 100$ generations) for all coculture
1116 experiments. For the Eco Eq 1 experiment, we mixed both S
1117 and L libraries in the same cultures along with wildtype L,
1118 using 4x 1L flasks. For the Eco Eq 2 experiments, S and L
1119 libraries were in separate cultures, both with wildtype S
1120 and L set at the appropriate frequency, with RB-Tn library
1121 frequency around 5 – 10% (Figure S6); cultures were propagated
1122 in 10x 50mL flasks. For the L in Maj and S in Maj 1
1123 experiments, we mixed wt L + S library and wt S + L library,
1124 respectively; cultures were propagated in 10x 50mL flasks.
1125 For the S in maj 2/3 experiments, we mixed wt S with S+L
1126 and L libraries respectively; cultures were propagated in 4x
1127 1L flasks.

1128 We measured the frequency of S/L in the population by plating
1129 and counting colonies at the end of a transfer on TM plates
1130 (tetrazolium maltose; 10g/L tryptone [Sigma T7293], 1g/L yeast
1131 extract [Sigma Y1625], 5g/L NaCl, 16g/L agar, 10g/L maltose,
1132 1mL/L 5% TTC [Sigma T8877]), where S appears as red colonies
1133 and L appears as white colonies, previously used in (35). We
1134 could also measure the frequency of cells from RB-Tn libraries
1135 by plating the cultures on LB/Kanamycin plates, as the transposon
1136 has a kanamycin resistance cassette (Figure S6). We diluted all
1137 cultures (at the end of a cycle) in DM0. Dilution rates varied
1138 over experiments: in Eco Eq 1, we diluted cultures by a factor
1139 of $2 * 10^{-5} mL^{-1}$ to plate on both TM and LB/Kan plates,
1140 in Eco Eq 2 we used dilution rates of $10^{-5} mL^{-1}$ and
1141 $10^{-4} mL^{-1}$ to plate on TM and LB/Kan plates respectively,
1142 in the L in Maj and S in Maj 1 experiments we used a
1143 $2 * 10^{-5} mL^{-1}$ dilution rate to plate on just TM plates,
1144 and in the S in maj 2/3 experiments we used dilution rates
1145 of $2 * 10^{-5} mL^{-1}$ and $2 * 10^{-4} mL^{-1}$ to plate on TM and
1146 LB/Kan plates respectively.
1147

1148 **S2.2.3 1:10 dilution.** We propagated cultures with a 1:10
1149 dilution, instead of the standard LTEE dilution rate of 1:100,
1150 to investigate the effect of a lengthened stationary phase relative
1151 to exponential phase. We used DM27.8 media (Davis Minimal
1152 Media with 27.8mg/L D-glucose), because the concentration of
1153 glucose would fall to 25mg/L after dilution. We used 1x 1L
1154 flask for each library culture (180mL media + 20mL culture),
1155 propagating the cultures for 8 days every 24hrs with $\log_2 10$
1156 generations per day. We pelleted and saved cultures every other
1157 day (0,2,4,6,8).

1158 **S2.2.4 Acetate exponential phase.** We sought to measure
1159 knockout fitness effects when the RB-Tn libraries were kept
1160 in acetate exponential phase, where we used DM2000-acetate
1161 (Davis Minimal Media with 2000mg/L Sodium Acetate) and
1162 grew the cultures in 1x 50mL flask. We first measured
1163 exponential growth rates for wt REL606, L, and S clones in

DM2000-acetate, which were approximately 0.08/hr, 0.12/hr, and 0.18/hr respectively. We also observed that all cultures were still in mid-exponential phase at $OD \sim 0.6$. So, if we started at initial OD_0 of 0.09, 0.03, 0.008 for REL606, L, and S respectively, the cultures would end up at $OD \sim 0.6$ after 24 hours. Thus, for each transfer, we would measure the actual OD for each culture (after 24hrs of growth) and transfer the appropriate volume of old culture to new 10mL DM2000-acetate such that the final concentration was the appropriate OD_0 . We recorded the number of generations for each cycle as $\log_2 OD_f / OD_0$. Due to the variable number of generations per transfer for each genetic background (owing to different growth rates), we collected samples at days 0,2,4,6,8 for REL606; 0,1,2,4,5,6 for L; 0,1,2,3,4,5 for S.

S2.2.5 Glucose exponential phase. We measured knockout fitness effects in glucose exponential phase with DM25 media in 1x 1L flask. We measured the length of DM25 exponential phase to be about 8.25 hrs for REL606, and 5.25 hrs for both S and L after a 1:100 dilution into new media. For the adaptation phase, we did two full 24hr cycles of growth in DM25, followed by one cycle of growth for ~ 8 hrs and ~ 5 hrs for REL606 and S/L, respectively. After the adaptation phase, we transferred cultures 1:100 into new DM25 media (warmed to 37C) four times, after 7.5-8hrs for REL606 and 4.5-5hrs for S and L. As DM25 media is quite dilute and thus OD measurements are relatively inaccurate, we estimated the number of cells that were transferred by plating the cultures on LB plates at a $2 * 10^{-5} mL^{-1}$ dilution rate and counting colonies, calculating the number of generations for that transfer as $\log_2 100 CFU_f / CFU_0$. We only ended up including the first two transfers of the REL606 library experiment (time points 0,1,2), as it was apparent from CFUs that the third transfer resulted in a large bottleneck owing to a smaller than expected population size before the transfer, likely because of slower than expected growth.

S2.3 DNA extraction, PCR, Sequencing. After the experiment was finished, pellets were pulled from the -80C freezer and genomic DNA was extracted with the Qiagen DNeasy tissue and blood extraction kit (cat no. 69504), eluted in double distilled water with typical yields around 50ng/ μ L. DNA barcodes were amplified from gDNA samples via PCR with Q5 Hot Start Polymerase (NEB, cat. no. M0493S); 50ul reactions were composed of 5 μ L PCR primers, 5 μ L gDNA, 10 μ L 5x buffer, 10 μ L GC enhancer, 1 μ L dNTPs, 0.5 μ L Q5 polymerase, 18.5 μ L water. We used custom dual-indexed primers that contained binding sites up- and down-stream of the barcode region, along with the necessary Illumina read/index binding sites; fwd primer (AATGAT ACGGCG ACCACC GAGATC TACACT CTTTCC CTACAC CAC GACGCT CTTCCG ATCT N_nXXXXXX GTCGAC CTGCAC CGTACG) where X stands for the custom forward 6bp index, and N_n is 1-4 random nucleotides, varying with the primer pair; rev primer (CAAGCA GAAGAC GGCATA CGAGAT XXXXXX GTGACT GGAGTT CAGACG TGTGCT CTTCCG ATCTGA TGTCCA CGAGGT CTCT) where X stands for standard Illumina 6bp IT index.

We used a different primer pair for each gDNA sample from a different experiment/replicate/time point, so that we could demultiplex the samples after sequencing. The PCR program was 4min at 95C, [30sec at 95C, 30sec at 55C, 30sec at 72C] x25 cycles, 5min at 72C. We verified that we had the correct PCR products via agarose gel electrophoresis. All PCR reactions were then pooled and cleaned with the Zymo DNA Clean and Concentrator kit (cat. no. D4013), and eluted in double distilled water. The final pooled sample was then sequenced on an Illumina HiSeq 4000 (50SR) at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

S3 Fitness inference pipeline

S3.1 Read counting and error correction. We first processed the raw (demultiplexed) sequencing reads using a previously developed Perl script (28, 29) that pulls out the barcode sequence by trimming regions corresponding to the sequencing primers and regions up/downstream of the barcode, as well as discarding reads that do not match the secondary sequencing index or have insufficiently high quality scores (MultiCodes.pl). Then, counts of unique barcodes are tabulated to get a table corresponding barcode sequence to counts. However, due to errors that arise during PCR and sequencing, some of the barcode reads acquire mutations that would prevent them from directly mapping to a transposon insertion location. Thus, we must correct for these sequencing errors by matching mutated barcodes to their parent, and merging the read counts together. The aforementioned Perl script identifies off-by-one barcode pairs; if the minority barcode (the one with fewer counts) unambiguously maps to a single majority barcode, the barcode counts are merged. To detect larger mutational distances between the derived and parent barcodes, we computed the Levenshtein (edit) distance between pairs of barcodes (as implemented in the Python C package Levenshtein (49)). Barcode read counts were merged if the edit distance was 4 or less, and if the minority barcode only mapped to one majority barcode at the minimum edit distance.

We then used previously acquired TnSeq data that maps the barcode identity to its transposon insertion location in order to identify which gene (if any) the barcoded transposon disrupted. Transposons that hit the first or last 5% of the gene sequence were excluded, as it is possible that these insertions do not result in disruption of production of the gene product. To ensure that barcodes at least begin their trajectories at a sufficiently high read count, if there were barcodes within a gene with low initial counts, $r_{i,0} < 80$, we summed the lowest (initial) count barcode into the next-lowest count barcode until $\min_i r_{i,0} \geq 80$. We restricted our analysis to genes that had ≥ 4 barcodes, allowing us to gain confidence that the measured knockout fitness is not dependent on rare fluctuations or secondary mutations. Additionally, some barcodes went extinct during the course of the experiment, either due to genetic drift or selection; if a barcode went extinct, i.e. has 0 counts from t_{ext} to T , we would trim all time points after, but not including, t_{ext} . We eliminated barcodes that go extinct after just one time point.

Library	# genes hit ≥ 3 times	# barcodes	% bc reads mapped
REL606	3,401	609,854	84%
6.5k S	3,382	522,253	84%
6.5k L	2,877	157,260	89%

Table S1. Summary of statistics of constructed RB-TnSeq libraries.

S3.2 Probabilistic model of read count trajectories and fitness inference. To infer the fitness of individual genotypes from BarSeq count data, we must first understand what frequency trajectories we would expect for a given fitness, and how technical noise (e.g. from sample preparation and sequencing) and genetic drift affect those trajectories. Consistent with previous work (7, 8, 12), we construct a maximum-likelihood estimator to infer fitness from trajectories of barcode read counts, using a deterministic approximation of frequency dynamics. On average, when the frequency of a lineage is sufficiently small $f_{t,i} \ll 1$, the frequency dynamics will exponentially grow/decay according to the genotype fitness, s , as well as the mean fitness of the population, \bar{x}_t (see section S3.4),

$$\langle f_{t,i} \rangle = f_{0,i} e^{(s - \bar{x}_t)t}$$

We measured the time in *generations*, which we measured for each time point in each experiment (see section S2.2). The reason we used a timescale of $1/\text{generation}$ instead of e.g. $1/\text{cycle}$ was to be able to better compare the magnitude of effects across experiments—e.g. the two exponential phase experiments had varying numbers of generations from cycle-to-cycle and between strains (due to differences in exponential growth rates). However, the fitness effects can be scaled by a factor of approximately 6.64 to get per-cycle fitness effects, at least in the 1:100 serial dilution experiments. The two sources of noise—genetic drift and measurement noise—both arise from counting processes, so the combined noise will follow $\text{var}(f_{t,i}) \propto \langle f_{t,i} \rangle$ (see section S3.3). To account for the inherent discreteness of counting sequencing reads—especially important to accurately model deleterious genotypes that quickly drop to low frequencies—we modeled the observed counts at time t (always measured in generations) of barcode i inserted in a given gene, $r_{i,t}$, as a negative binomial random variable,

$$r_{i,t} | s, f_{0,i} \sim \text{NB}(\mu_{i,t}, c_t) \quad (1)$$

$$\langle r_{i,t} \rangle = \mu_{i,t} \quad (2)$$

$$\text{var}(r_{i,t}) = c_t \langle r_{i,t} \rangle \quad (3)$$

$$\mu_{i,t} = R_t f_{0,i} e^{(s - \bar{x}_t)t} \quad (4)$$

Where R_t is the total number of counts, and c_t is the measured variance parameter. The final likelihood for the fitness, s , of a given gene knockout is obtained by numerically integrating over $f_{0,i}$ ('integrated likelihood' with a flat prior)—incorporating the uncertainty in the intercept nuisance parameters into the fitness estimate and turning the problem into a one-dimensional maximum likelihood—and then combining the likelihoods of all barcodes inserted into the gene,

$$P(r_i | s, f_{0,i}) = \prod_t \frac{\Gamma(r_{i,t} + \frac{\mu_{i,t}}{c_t - 1})}{\Gamma(\frac{\mu_{i,t}}{c_t - 1}) \Gamma(r_{i,t} + 1)} \frac{(c_t - 1)^{r_{i,t}}}{c_t^{r_{i,t} + \frac{\mu_{i,t}}{c_t - 1}}} \quad (5)$$

$$\mathcal{L}(s | r) = \prod_i \int df_{0,i} P(r_i | s, f_{0,i}) \quad (6)$$

The point estimate of the knockout fitness, \hat{s} , is then numerically computed as the maximum likelihood, and the standard error is approximated as the inverse, square-root observed information,

$$\hat{s} = \underset{s}{\text{argmax}} \log \mathcal{L}(s | r) \quad (7)$$

$$\text{std } \hat{s} = 1 / \sqrt{-\partial_s^2 \log \mathcal{L}(s | r) |_{\hat{s}}} \quad (8)$$

We ran biological replicates for all experiments reported here; to obtain combined genotype fitness estimates across replicates we simply multiplied the likelihoods together, repeating the maximum likelihood procedure.

As the majority of barcoded knockouts are neutral or nearly so ($s \approx 0$), we must have a method to distinguish between likely neutral and selected knockout mutations; this can be accomplished by computing a p-value under the null hypothesis $s = 0$. For ease of computation and generality we compute the p-value as the posterior probability that the likelihood ratio between null and alternative hypotheses is greater than 1, i.e. the probability that the data more strongly support the null hypothesis over the alternative,

$$p = P_{s|r} \left(\frac{\mathcal{L}(0|r)}{\mathcal{L}(s|r)} > 1 \right)$$

$$P(s|r) \propto \mathcal{L}(s|r)$$

This convenient definition has been shown to be equivalent to the frequentist definition of the p-value using a likelihood ratio test statistic (50, 51), and does not require asymptotic approximations. We used the standard method of Benjamini & Hochberg to control for the false discovery rate at $\alpha = 0.05$.

S3.3 Estimation of error parameters. In order to estimate fitness of individual genotypes from BarSeq data, we must first obtain an estimate of the error parameters for each time point in the experiments. There are two distinct sources of noise in our BarSeq measurements—measurement (technical) noise, arising from library preparation and sequencing error, which is uncorrelated in time, and variance due to genetic

1347 drift, which accumulates over time. Both sources of noise are
 1348 count processes, where the variance of barcode population
 1349 frequencies will be proportional to the mean,

$$\langle f_{i,t} \rangle = \frac{\langle r_{i,t} \rangle}{R_t} \propto \text{var}(f_{i,t})$$

1350 In order to eliminate the dependence of the variance on the
 1351 mean, we apply a variance-stabilizing transformation,

$$\phi_{i,t} \equiv \sqrt{f_{i,t}}$$

1352 The variance of barcode frequencies of neutral lineages over
 1353 two time points will then depend on the variance that has ac-
 1354 cumulated due to genetic drift, as well as the technical noise
 1355 at the sampled time points. If there are sufficiently many read
 1356 counts/individuals such that the central limit theorem applies,
 1357 the variances will simply be additive,

$$\kappa_{j,k} \equiv \text{var}(\phi_{i,j} - \phi_{i,k}) = \zeta_j + \zeta_k + \frac{|j-k|}{4N_e} \quad (9)$$

1358 Where ζ_t is the technical noise at time point t , N_e is the ef-
 1359 fective population size, and $|j-k|$ is the number of transfers
 1360 performed between times j and k . The above equation de-
 1361 fines a set of linear equations, with ζ_t and N_e as unknown
 1362 parameters.

1363 We can measure $\kappa_{j,k}$ for all possible combinations of t_j and
 1364 t_k given large enough set of neutral barcodes. Our RB-TnSeq
 1365 libraries have a large number of transposons that were in-
 1366 serted into intergenic regions, the vast majority of which pre-
 1367 sumably have no fitness effect; thus, we use these intergenic
 1368 barcodes as our set of putatively neutral barcodes. We con-
 1369 firmed that our measured $\kappa_{j,k}$ did not systematically vary
 1370 as a function of r_j (Figure S1), indicating that the expected
 1371 mean-variance relationship, $\langle f_{i,t} \rangle \propto \text{var}(f_{i,t})$, is consistent
 1372 with our data.

1373 We only included intergenic barcodes that satisfy $50 < r_{i,t} <$
 1374 500 , as our computation depends on having sufficiently many
 1375 counts such that the central limit theorem applies, and bar-
 1376 codes at a higher frequency are more likely to have acquired
 1377 secondary mutations and be impacted by selection. In order
 1378 to further guard against the effects of potential 'outlier'
 1379 barcodes (those with non-neutral fitnesses), we compute vari-
 1380 ance estimates, $\hat{\kappa}_{j,k}$, with a more robust measurement of vari-
 1381 ability, the median absolute deviation (MAD),

$$\psi_{i,j,k} \equiv \phi_{i,j} - \phi_{i,k} \quad (10)$$

$$\text{MAD}_{j,k} = \text{med}_i |\psi_{i,j,k} - \text{med}_i \psi_{i,j,k}| \quad (11)$$

$$\hat{\kappa}_{j,k} = \left(\frac{\text{MAD}_{j,k}}{0.67449} \right)^2 \quad (12)$$

1382 We resampled barcodes with replacement (standard boot-
 1383 strapping) 500 times to compute the relative errors on the
 1384 $\hat{\kappa}_{j,k}$ measurements. To decompose variability into the cor-
 1385 related ($1/N_e$) and uncorrelated (ζ_t) components, we numeri-
 1386 cally minimized squared error of the expected relationship



Fig. S1. The measured noise parameter $\kappa_{j,k}$ is consistently approximately constant as a function of initial number of barcode reads. Data is from S/L/REL606 monoculture experiments, replicate 1. Curves are smoothed with a moving average, ± 2 reads.

(eq. 9) between the noise parameters and the measured $\hat{\kappa}_{j,k}$,
 with inverse variance weighting,

$$\zeta, N_e = \underset{\zeta, N_e}{\text{argmin}} \sum_{j,k} \frac{\left(\zeta_j + \zeta_k + \frac{|j-k|}{4N_e} - \hat{\kappa}_{j,k} \right)^2}{\text{var}(\hat{\kappa}_{j,k})}$$

1387 We subjected the minimization to the constraint that $\zeta_t \geq$
 1388 $\frac{1}{4R_t}$, i.e. technical noise must be at least as large as vari-
 1389 ance due to sampling. After converting the variance paramet-
 1390 ers from frequencies back to read counts, the total marginal
 1391 variance parameter at a single time point is,
 1392
 1393

$$\hat{c}_t = (4\zeta_t + 1/N_e)R_t$$

1394 The number of intergenic barcodes included varies across
 1395 RB-TnSeq libraries, experiments, and time points, but ap-
 1396 proximately on the order of $\sim 10^4$ intergenic barcodes are
 1397 used to estimate the variance parameters. The errors on the
 1398 estimated \hat{c}_t are generally small ($\lesssim 1\%$), so the point estimate
 1399 \hat{c}_t was directly used for all downstream inferences.

S3.4 Estimation of mean fitness dynamics. As benefi-
 cial mutations increase in frequency, and deleterious muta-
 tions decrease, the mean fitness of the population changes
 over time, impacting the rate of frequency change of all gen-
 otypes in the population. To estimate the mean fitness dynam-
 ics for each experiment, we can track the dynamics of neutral
 genotypes, again using the large set of intergenic barcodes.
 We obtain an estimate of the mean fitness between times 0
 and t by simply taking the negative log slope over many bar-
 codes,

$$\hat{x}_{i,t} = -\frac{1}{t} \left(\log\left(\frac{r_t}{R_t}\right) - \log\left(\frac{r_0}{R_0}\right) \right)$$

As detailed in the previous section, it is advantageous to use
 robust forms of estimation to guard against the presence of

1412 outliers. Groups of ~ 100 randomly selected intergenic bar- 1461
 1413 codes with $r_{i,t} < 500$ were summed together to create "super- 1462
 1414 barcodes", in order to improve individual estimates. The 1463
 1415 mean fitness $\hat{x}_{i,t}$ was estimated for each super-barcode sep- 1464
 1416 arately, and then the final estimate \hat{x}_t was obtained by tak- 1465
 1417 ing the median over all super-barcodes. The standard error 1466
 1418 was estimated via the median absolute deviation between all 1467
 1419 super-barcodes, analogous to equations 11-12. Again, the 1468
 1420 point estimate \hat{x}_t is used for all downstream analyses, as 1469
 1421 mean fitness error was consistently small. 1470

1422 **S3.5 Identification of putative outlier barcodes.** We ob- 1465
 1423 served that some barcodes had trajectories that noticeably dif- 1466
 1424 fered from the rest of the barcodes within the genotype, likely 1467
 1425 caused by the presence of secondary (selected) mutations that 1468
 1426 arose elsewhere in the genome or rare frequency fluctuations. 1469
 1427 We observed outlier barcodes with both beneficial and dele- 1470
 1428 terious trajectories relative to the rest of the barcodes within 1471
 1429 the genotype. Problematically, some of these outlier bar-
 1430 codes were at high abundance relative to the other barcodes in
 1431 the genotype, thus dominating the genotype fitness estimate.
 1432 This necessitated a need to either accommodate outliers in
 1433 our fitness estimation procedure or detect and reject outliers.
 1434 We found that a number of robust estimators that we explored
 1435 (e.g. maximum median/trimmed likelihood) had unreason-
 1436 ably high variance in fitness given our data ($\text{std } \hat{s} \gtrsim \hat{s}$). Thus,
 1437 we opted to use a method to detect and reject outlier barcodes
 1438 within genotypes. We based our outlier detection method on
 1439 the resistant diagnostic RD_i introduced by Rousseeuw and
 1440 Leroy (1987) (52), a high-breakdown measure of statistical
 1441 deviation.

1442 For every genotype with at $n_{bc} \geq 4$ unique barcodes, we
 1443 computed a fitness estimate for each barcode, \hat{s}_i , via max-
 1444 imum likelihood (eqs. 5-8). We then used a resampling
 1445 approach to randomly sample 200 different combinations of
 1446 $n_r = \lceil n_{bc}/2 \rceil$ barcodes, where samples are labeled J . To get
 1447 an estimate of the 'typical' fitness, $\hat{s}_{J,typ}$, of the barcodes
 1448 within a gene, we either take the weighted median ($n_r < 10$)
 1449 or weighted trimmed mean ($n_r \geq 10$, trim 30% off each tail)
 1450 of the resampled barcode fitnesses, where in both cases, sam-
 1451 ples are weighted by their inverse variance, $w_i = 1/(\text{var } \hat{s}_i)$.
 1452 The weighted median is used for low number of samples,
 1453 while the trimmed weighted mean is used for high number
 1454 of samples, because the trimmed weighted mean generally
 1455 has lower sampling variance when the number of samples re-
 1456 maining after trimming is sufficiently large. To compare the
 1457 strength of evidence for a fitness of \hat{s}_i or $\hat{s}_{J,typ}$ for barcode
 1458 i , we compute the likelihood ratio,

$$\text{LR}_{J,i} = \log \frac{\mathcal{L}_i(\hat{s}_i | r_i)}{\mathcal{L}_i(\hat{s}_{J,typ} | r_i)}$$

1459 The deviation of barcode i from the rest of the barcodes in
 1460 the genotype is then,

$$u_i = \max_J \frac{\text{LR}_{J,i}}{\text{med}_i \text{LR}_{J,i}}$$

The final resistant diagnostic is finally calculated as a stan-
 dardized version of u_i ,

$$RD_i = \frac{u_i}{\text{med}_i u_i}$$

If $RD_i > \text{cutoff}$, then barcode i is considered an outlier and
 thrown away.

S3.5.1 Simulations. To determine an appropriate cutoff
 value, we performed simulations of the data generating
 process, and calculated the RD for each barcode within
 a simulated gene using the above method. Specifically,
 we simulated trajectories of lineage frequencies with $s \in$
 $\{-0.02, 0, 0.02\} \text{ gen}^{-1}$ with the standard diffusion approx-
 imation, assuming $f \ll 1$,

$$\begin{aligned} \partial_t f &= sf + \sqrt{\frac{f}{N_e}} \eta(t) \\ \langle \eta(t) \rangle &= 0 \\ \langle \eta(t) \eta(t') \rangle &= \delta(t - t') \end{aligned}$$

We 'observed' trajectories at the end of each 'day' (≈ 6.64
 gen) for 4 days, and added measurement noise,

$$\begin{aligned} \phi_t &\equiv \sqrt{f_t} \\ \phi_t^{obs} | \phi_t &\sim \mathcal{N}(\phi_t, \zeta) \end{aligned}$$

We used $N_e = 10^8 \text{ day}$ and $\zeta = 2 * 10^{-8}$. We then grouped
 20 simulated lineages together into a 'gene' (approximate
 median number of barcodes per gene in our libraries), with
 $n \in \{1, 2, 3\}$ selected lineages (of the same sign), and the rest
 as neutral lineages. After calculating the RD for each sim-
 ulated gene, we calculated the true positive/negative rate for
 calling a lineage as an outlier for a given threshold (Figure
 S2).

We can see that the method can sensitively detect relatively
 small, $\sim 2\%$, differences in fitness, while minimizing the
 number of neutral barcodes that are incorrectly thrown away.
 True positive rate decreases somewhat if there are multiple
 outlier barcodes within a gene, but the difference appears to
 be minimal, as expected from the construction of the RD as
 a high-breakdown deviance statistic. From the simulations,
 we chose a cutoff of 6, which only falsely throws out $\sim 5\%$
 of neutral lineages, while detecting $\sim 85 - 95\%$ of outliers.
 This threshold also seems to empirically work with our data,
 detecting at least the most obvious outliers (see e.g. Figure
 S3).

**S3.6 Consequences of potential barcode frequency
 biases.** One major assumption of the above analyses is that
 the frequency of barcodes from BarSeq data represents an *un-*
biased estimate of the actual frequency of barcoded cells in
 the population. While we expect this assumption to generally
 hold, there are two major ways that this assumption could be
 violated: (1) if barcodes are differentially amplified due to
 e.g. differences in GC content, and (2) if genomic regions

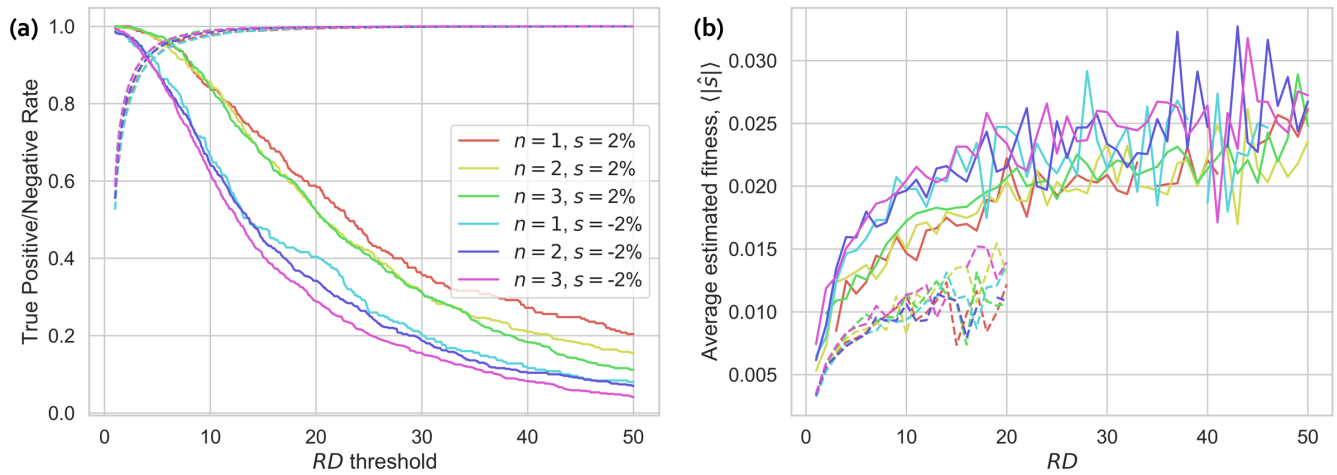


Fig. S2. (a) Detection of selected, outlier barcodes in otherwise neutral genes. Dotted lines are the true negative rate, solid lines are the true positive rate. (b) Average inferred fitness (ie apparent fitness, differing from the true fitness by fluctuations) of barcodes with different RDs. Dotted lines are from neutral barcodes, solid lines are outlier barcodes. ‘Neutral’ barcodes with $RD \approx 6$ have sufficiently large fluctuations to have trajectories that appear to have a 1% deviation from neutrality.

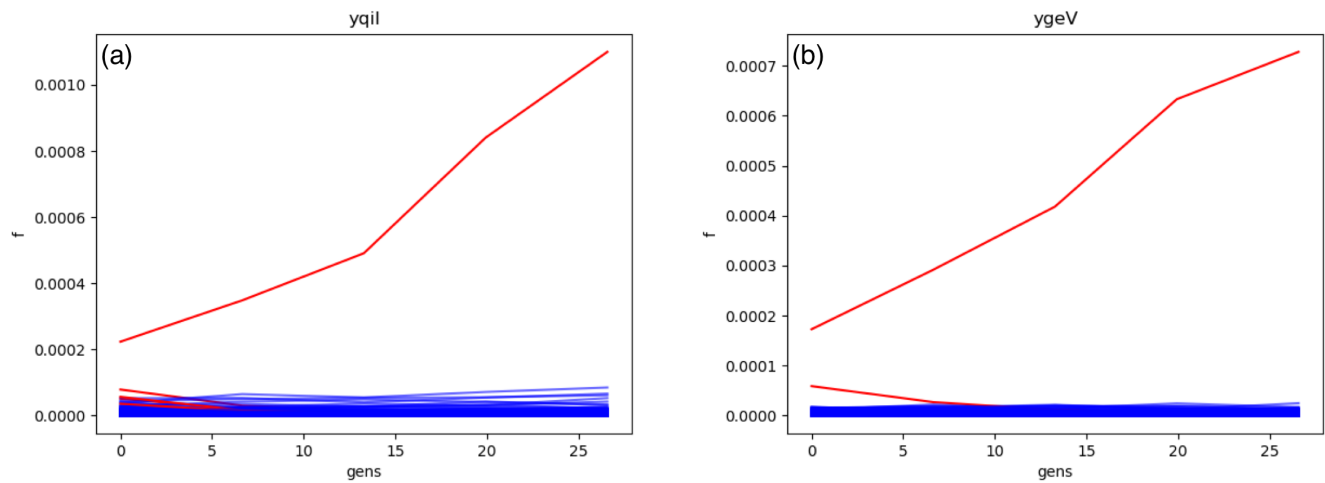


Fig. S3. Examples of high-abundance outlier barcodes detected in otherwise neutral genotypes. Red barcodes were called as outliers. Examples taken from an experiment with the 6.5k S library in co-culture with L at the equilibrium frequency.

1502 near the chromosomal origin of replication are present at a
 1503 higher copy number due to fast growth. Both types of biases
 1504 have been observed in some previous RB-TnSeq experiments
 1505 (28, 29). We can check for the presence of frequency biases
 1506 by comparing the inferred value of the error parameter κ_t
 1507 (see section S3.3) for barcodes with different GC contents
 1508 and across genomic positions, as biases in frequency mea-
 1509 surements will change the apparent strength of genetic drift.
 1510 We see that κ_t generally does not change across these con-
 1511 ditions (Figure S4), and thus the aforementioned sources of
 1512 frequency biases do not seem to be particularly prevalent or
 1513 strong in our system.

1514 Of course, other unknown sources of frequency bias could
 1515 be present, or too weak to detect; but, under our inference
 1516 pipeline, biases in frequency would only affect the variance
 1517 of inferred s , not its expected value, as long as the bias across
 1518 time points remains constant. We can see this by considering
 1519 the deterministic (mean) dynamics of mutant frequencies f
 1520 in a population with m genotypes,

$$f_i(t) = \frac{f_{i,0}e^{s_i t}}{\sum_j^m f_{j,0}e^{s_j t}}$$

We could then include a strain-specific, constant multiplica- 1521
 tive bias parameter, γ_i . The observed frequencies would then 1522
 follow, 1523

$$f_i(t) = \frac{\gamma_i f_{i,0} e^{s_i t}}{\sum_j^m \gamma_j f_{j,0} e^{s_j t}}$$

By observing these biased frequencies instead of the actual 1524
 frequencies, we would infer s_i and $\gamma_i f_{i,0}$, therefore only bi- 1525
 asing the nuisance intercept parameter. 1526

S4 Analysis 1527

S4.1 Correlation of fitness effects across environ- 1528
ments. To compute the correlation of knockout fitness ef- 1529
 fects across environments for a given genetic background 1530

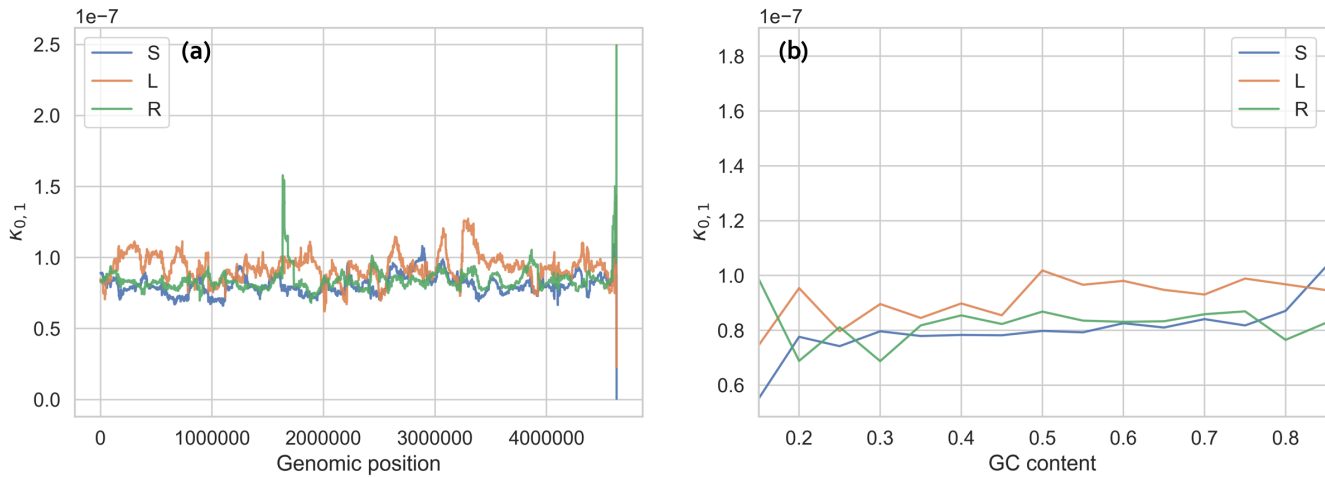


Fig. S4. The measured noise parameter $\kappa_{j,k}$ does not vary systematically over (a) genomic position, or (b) barcode GC content, indicating that these factors do not measurably bias barcode frequency measurements. Data is from S/L/REL606 monoculture (1) experiments, replicate 1.

(main text Figure 3), we first removed genes with noisy fitness effects ($\sigma_s > 1\%$), then calculated the weighted Pearson correlation coefficient, where genes are labeled k and environments are labeled i, j ,

$$w_k = 1/(\text{var } \hat{s}_{i,k} + \text{var } \hat{s}_{j,k}) \quad (13)$$

$$\mu(x) = \frac{\sum_k w_k x_k}{\sum_k w_k} \quad (14)$$

$$\text{wcov}(x, y; w) = \frac{\sum_k w_k (x_k - \mu(x))(y_k - \mu(y))}{\sum_k w_k} \quad (15)$$

$$\rho_{i,j} = \frac{\text{wcov}(\hat{s}_i, \hat{s}_j; w)}{\sqrt{\text{wcov}(\hat{s}_i, \hat{s}_i; w)\text{wcov}(\hat{s}_j, \hat{s}_j; w)}} \quad (16)$$

We then performed hierarchical clustering using Ward's method across environments for each genetic background, with $1 - \rho_{i,j}$ as the distance metric. Environment pairs with $\rho_{i,j} < 0$ are set to 0 for the purposes of clustering, as there were few negative correlations, and all were small.

We used a bootstrapping procedure to estimate the statistical support for each cluster of environments. Using only the intersection of genes that passed across all environments, we performed standard resampling of genes with replacement, and then repeated the correlation measurement of knockout fitness values for each pair of environments. Then we repeated the hierarchical clustering and compared each branching of the original tree to the bootstrapped tree using the method of (53). We repeated the resampling procedure 5000 times for each genetic background and reported the average support for each clade.

S4.2 Network of gene-by-gene correlations. To investigate potential relationships between genes in the different strain investigated in our work, we sought to quantify the degree of correlation of fitness measurements across all environments between every pair of genes, a quantity that has previously been referred to as cofitness (29). Highly correlated fitness measurements may indicate that genes are connected

via gene regulation. In order to account for the fact that the measurement error in fitness measurements varies between genes and environments, we computed the cofitness of every pair of genes i, j as the weighted Pearson correlation coefficient, where environments are labeled k , analogous to equations 13-16. We excluded genes that were not called as significantly non-neutral in at least one experiment, and genes with successful fitness measurements in < 4 experiments.

The vast majority of non-zero correlations are likely generated by chance, due to the relatively small number of environments where fitness is measured. Therefore, for each pair of genes, we generated a null cofitness distribution through a resampling procedure performed 300 times, by (1) randomly permuting the fitness assignments for both genes, (2) resampling each fitness value such that $\hat{s}_{boot} \sim \mathcal{N}(\hat{s}, \text{std } \hat{s})$ ("parametric bootstrapping"), and (3) recalculating cofitness via equations 13-16. We then compared the measured cofitness to the null distribution to generate a 1-sided p-value. After correcting the set of p-values with a Benjamini-Hochberg FDR correction, we considered gene pairs to be significantly correlated at $\alpha = 0.05$, effectively drawing an edge between the two genes in the cofitness network.

After identifying statistically significant correlations between genes across environments, we sought to cluster genes into communities, without considering the magnitude or sign of the cofitness values. We used the 'Fluid Communities' algorithm (42), as implemented in the `networkx` python package (54), because of the flexibility of the algorithm, and the resulting communities had the highest modularity of all community-finding algorithms we explored. As the fluid communities algorithm is initialized stochastically, and requires pre-specifying k communities, we ran the algorithm on our data across varying community sizes, $k \in [4, 20]$, with 200 replicates for each k (Figure S16). We then picked the communities with the highest modularity for each genetic background. For the purposes of community finding, we treated all significant edges as the same, without considering the actual cofitness value of the edge. All community

1596 sets found had $modularity > 0$, indicating that genes were
1597 more tightly connected within their community compared to
1598 between communities.

1599 Standard gene ontology enrichment analysis was performed
1600 on each community in each genetic background with the
1601 `goatools` python package (55), using Fisher's exact test to
1602 find significantly over-represented annotations in a gene set,
1603 with an FDR correction and $\alpha = 0.05$.

1604 We sought to check if variance in fitness across environments
1605 for any given knockout could predict if two genes would stay
1606 in the same cluster across genetic backgrounds, as a control
1607 for the observed correlation with EcoliNet score. We average
1608 fitness variance across environments over the two knockouts
1609 of interest, referring to the quantity as $\langle var(s) \rangle$. We fit a lo-
1610 gistic model with normalized EcoliNet score of the gene pair,
1611 $n_{score} \equiv score/stdscore$ and $nvar \equiv \langle var(s) \rangle / std \langle var(s) \rangle$
1612 as the predictors (standard deviation is taken over all knock-
1613 out pairs), and the probability that the two genes are together
1614 in strain 2, if they were together in strain 1 as the response
1615 variable, $\log p_i / (1 - p_i) = n_{score} \beta_{score} + nvar \beta_{var} + \beta_0 +$
1616 ϵ_i . The results are shown in Figure S22.

1617 It is known that community detection algorithms can have
1618 potential surfaces with large plateaus without a clear max-
1619 imum, i.e. can give many solutions with similar modularity
1620 but different groupings (56). We wanted to see if the observed
1621 (mostly) "random reassortment" of genes among clusters be-
1622 tween genetic backgrounds could be explained by this effect.
1623 Thus, we compared the optimal partition of each background
1624 to the 100 next-best partitions across all backgrounds (Figure
1625 S17). For each suboptimal partition, we asked if two genes
1626 were in the same cluster in the optimal partition, what is the
1627 probability that they are also in the same cluster in the sub-
1628 optimal partition. We see that if we compare partitions in
1629 the same genetic background, this probability is around 40%,
1630 while it is around 10% when comparing partitions across
1631 background. This suggests that different reasonable parti-
1632 tions of the cofitness networks are much more similar within
1633 genetic backgrounds than between backgrounds. We also
1634 re-ordered the genes of the cofitness network such that they
1635 followed the ordering of another genetic background's opti-
1636 mal partition (Figure 4B). It is apparent that replotting the
1637 cofitness matrix using another genetic background's cluster-
1638 ing does not produce noticeable structure. Together, these
1639 results suggest that while different reasonable partitions can
1640 give slightly different clusters, the observed reassortment of
1641 knockout fitness correlations among backgrounds cannot just
1642 be explained by failures of the community detection algo-
1643 rithm. We also investigated the extent to which the struc-
1644 ture of our cofitness networks was driven by measurement
1645 noise (Figure S18). We leveraged the fact that we had at least
1646 two biological replicates per experiment, and computed new
1647 cofitness networks (in the same manner as described above),
1648 only using either biological replicate "1" or "2". We can see
1649 that even when the data is independently split, the cofitness
1650 networks within a genetic background are more similar than
1651 between backgrounds.

S4.3 Genome evolution. We sought to understand if
1652 knockout fitness measurements could predict the probability
1653 that a gene would mutate in the LTEE. To that end, we down-
1654 loaded clonal sequencing data from Tenaillon et al. (2016)
1655 (46), where the authors isolated and sequenced clones from
1656 a number of time points across all 12 lines of the LTEE, and
1657 identified mutations relative to the REL606 ancestor. We ex-
1658 cluded synonymous SNPs from our analysis. A representa-
1659 tion of the raw data can be found in Figure S26.
1660

We then sought to understand if knockout fitness effects can
1661 predict if a mutation will appear in a gene in the Tenaillon
1662 et al. dataset, as a proxy for establishment. For REL606,
1663 classified a gene as mutated if a mutation appeared in one of
1664 the 12 LTEE lines (excluding mutator populations). For S and
1665 L, we classified genes as mutated only if they were present
1666 in the appropriate sublineage, i.e. in REL11830, REL11036
1667 or REL11831, REL11035 for S and L respectively. We also
1668 excluded mutations that were already present in our S and
1669 L clones, which we determined from clonal sequencing data
1670 from Plucain et al. (2014) (35). We then fit a logistic model
1671 with knockout fitness effect as the predictor variable and gene
1672 mutated status (between time points) as the response variable,
1673

$$\log p_{est,i} / (1 - p_{est,i}) = \pm \tilde{s}_i \beta_{est\pm} + \beta_0 + \epsilon_i$$

We fit two different coefficients for beneficial and deleteri-
1674 ous mutations in each environment, β_{est+} and β_{est-} respec-
1675 tively. We only include genes that are putatively neutral, i.e.
1676 $|s| < 0.005$ and not called as significantly non-neutral, along
1677 with genes that are either significantly beneficial or deleteri-
1678 ous, all at significance level $\alpha = 0.05$. We normalized the
1679 fitness values by the median value of the non-neutral genes,
1680 i.e.
1681

$$\tilde{s}_i = \frac{s_i}{\text{med}_{i \notin \text{neutral}} s_i} \quad (17)$$

We use the logistic model implementation in the
1682 `statsmodels` python package (57). We used the
1683 standard method of Benjamini & Hochberg to control for
1684 the false discovery rate, pooling all tests across beneficial
1685 and deleterious coefficients. To test if there is a significant
1686 difference between REL606 logit slopes at 0-5k and 5-20k,
1687 we employed a permutation test. To construct a null distri-
1688 bution of the difference in slopes, for each gene we shuffled
1689 whether it 'established' (0 or 1) between 0-5k and 5-20k and
1690 recomputed the regression coefficients 1000 times, recording
1691 the difference. We then compared the actual difference in
1692 coefficients to the null distribution to get p-values.
1693

S4.4 Changes in gene expression. We used a microar-
1694 ray gene expression dataset previously reported by Le Gac et
1695 al. (2012) (34) to compare to our knockout fitness measure-
1696 ments, downloaded from the NCBI Gene Expression Omnibus
1697 (58), importing data with GEOquery (59). We primar-
1698 ily used the GEO2R tool to process the raw microarray
1699 data along with the R package `limma` (60, 61). After ap-
1700 plying a \log_2 transform to the data, we ensured that all col-
1701

1702 lected samples had approximately the same intensity distri-
1703 butions by performing a quantile normalization. Then, pool-
1704 ing all replicates within a strain, we fit a linear model to our
1705 data to determine the relative log-fold change in expression
1706 between different strains, taking into account the measured
1707 mean-variance relationship. A representation of the raw data
1708 can be found in Figure S25.

1709 We then fit a linear model investigate if there was a relation-
1710 ship between fitness measured in a given environment, s_i ,
1711 and log-change in gene expression between evolutionary time
1712 points ΔE_i , such that

$$\Delta E_i = \pm \tilde{s}_i \beta_{exp\pm} + \beta_0 + \epsilon_i$$

1713 Similar to the gene establishment model, we fit two differ-
1714 ent coefficients for beneficial and deleterious mutations in
1715 each environment, β_{exp+} and β_{exp-} respectively. We only
1716 include genes that are putatively neutral, i.e. $|s| < 0.005$ and
1717 not called as significantly non-neutral, along with genes that
1718 are either significantly beneficial or deleterious, all at signif-
1719 icance level $\alpha = 0.05$. We normalized the fitness values by
1720 the median value of the non-neutral genes, in the same man-
1721 ner as equation 17. We fit the model with weighted least
1722 squares, as implemented in the `statsmodels` python pack-
1723 age (57), with weights $w_i \propto 1/\text{var} \Delta E_i$, to incorporate the
1724 fact that there are different levels of measurement error in
1725 the log-fold change expression for each gene. We used the
1726 standard method of Benjamini & Hochberg to control for the
1727 false discovery rate, pooling all tests across beneficial and
1728 deleterious coefficients.

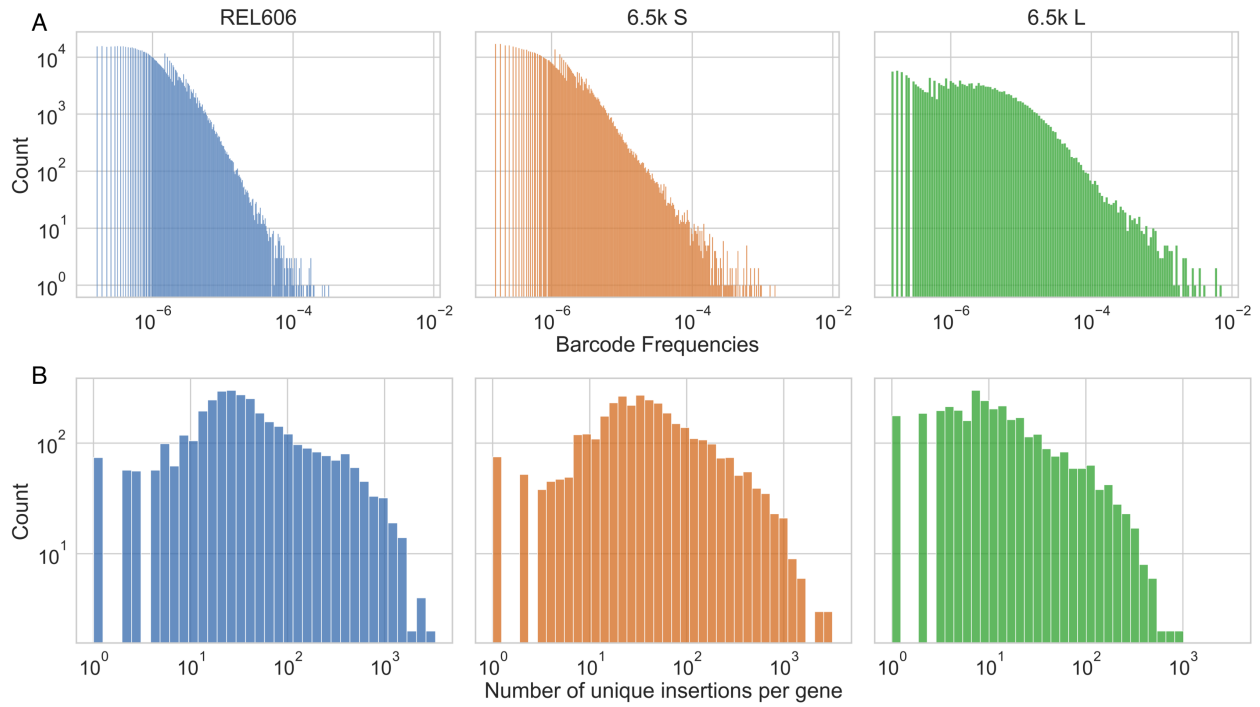


Fig. S5. Statistics of RB-TnSeq libraries, **(A)** initial distribution of barcode frequencies in library populations, and **(B)** distribution of number of unique barcoded transposon insertions into each gene (cds).

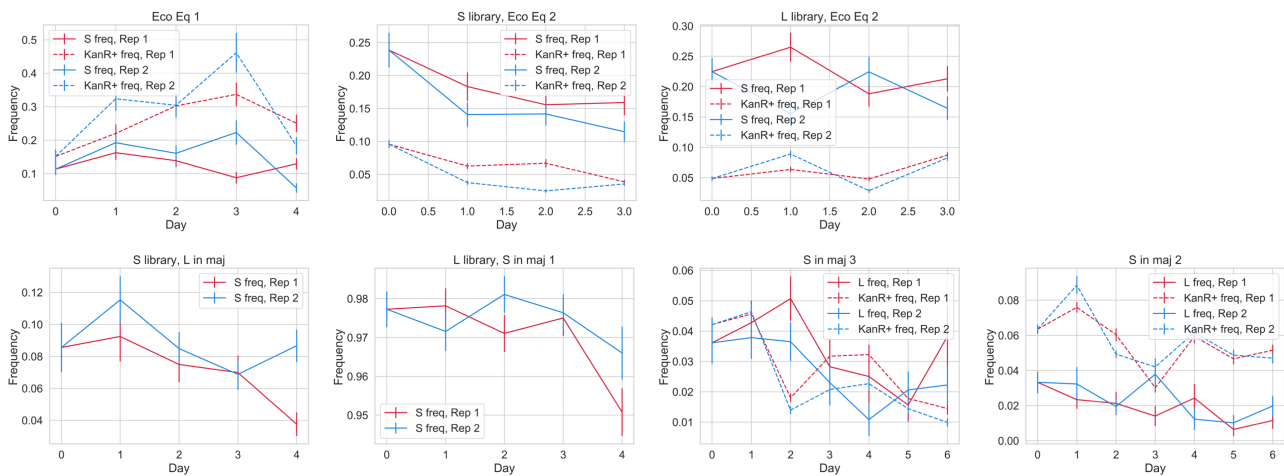


Fig. S6. Frequency trajectories of mixed culture experiments from CFUs. For each coculture experiment, we diluted and plated cultures on both TM plates (S/L indicator plates) and LB/Kan plates (pulls out cells from the RB-TnSeq libraries). We didn't plate experiments "S/L in maj (1)" on LB/Kan plates because we only cocultured wt S/L with L/S RB-TnSeq libraries respectively. Please note that each subplot is on a different scale.

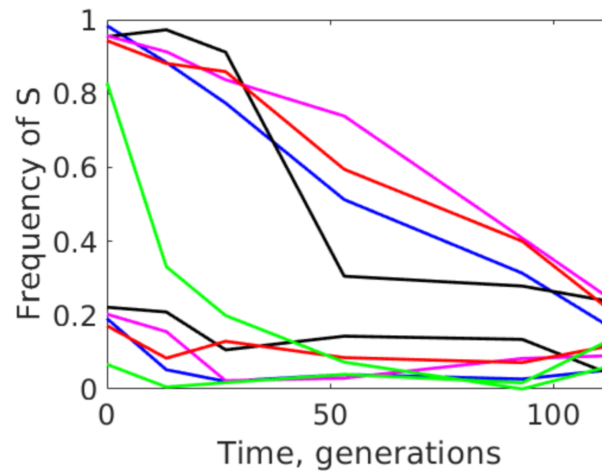


Fig. S7. Measured S/L frequency dependent fitness and ecological equilibrium via CFUs on TM plates (S/L indicator plates). Cocultures of S and L wt clones were propagated in standard LTEE conditions.

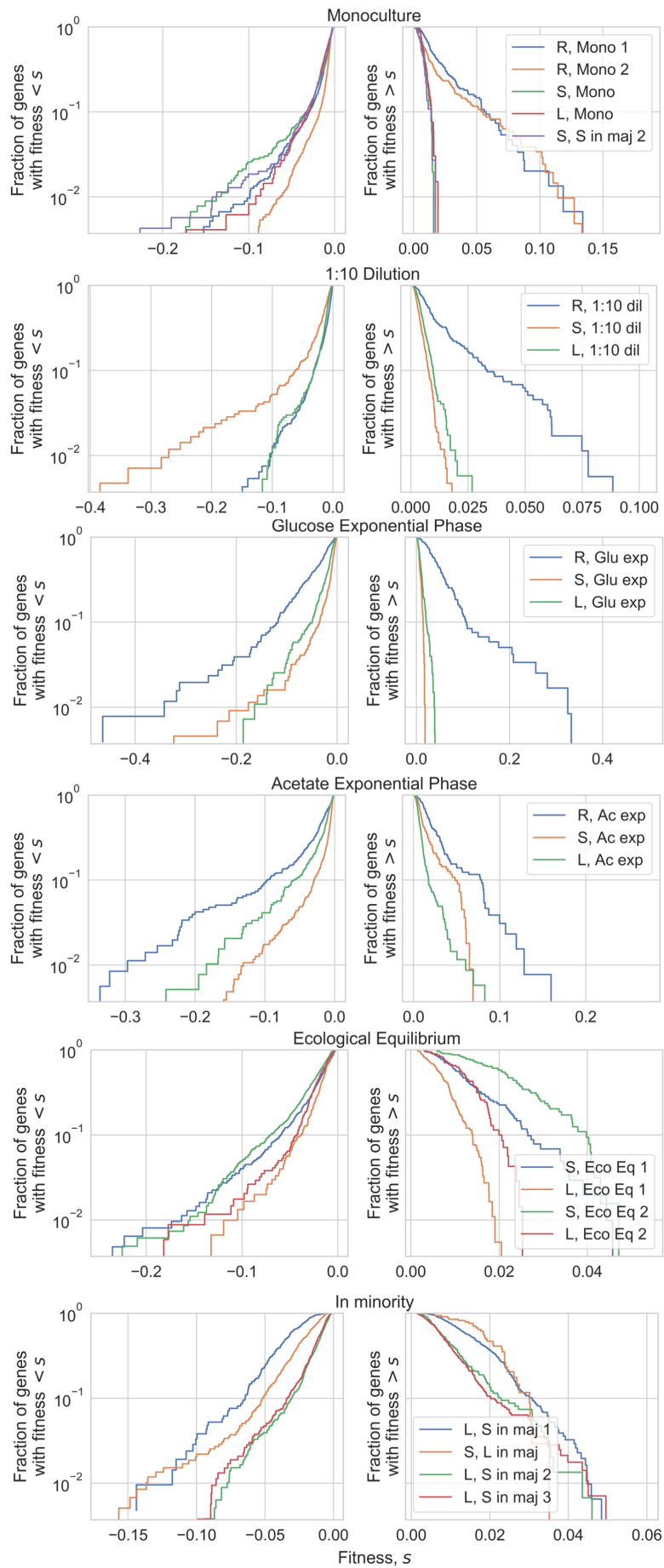


Fig. S8. All measured DFEs across experiments, arranged by environment.

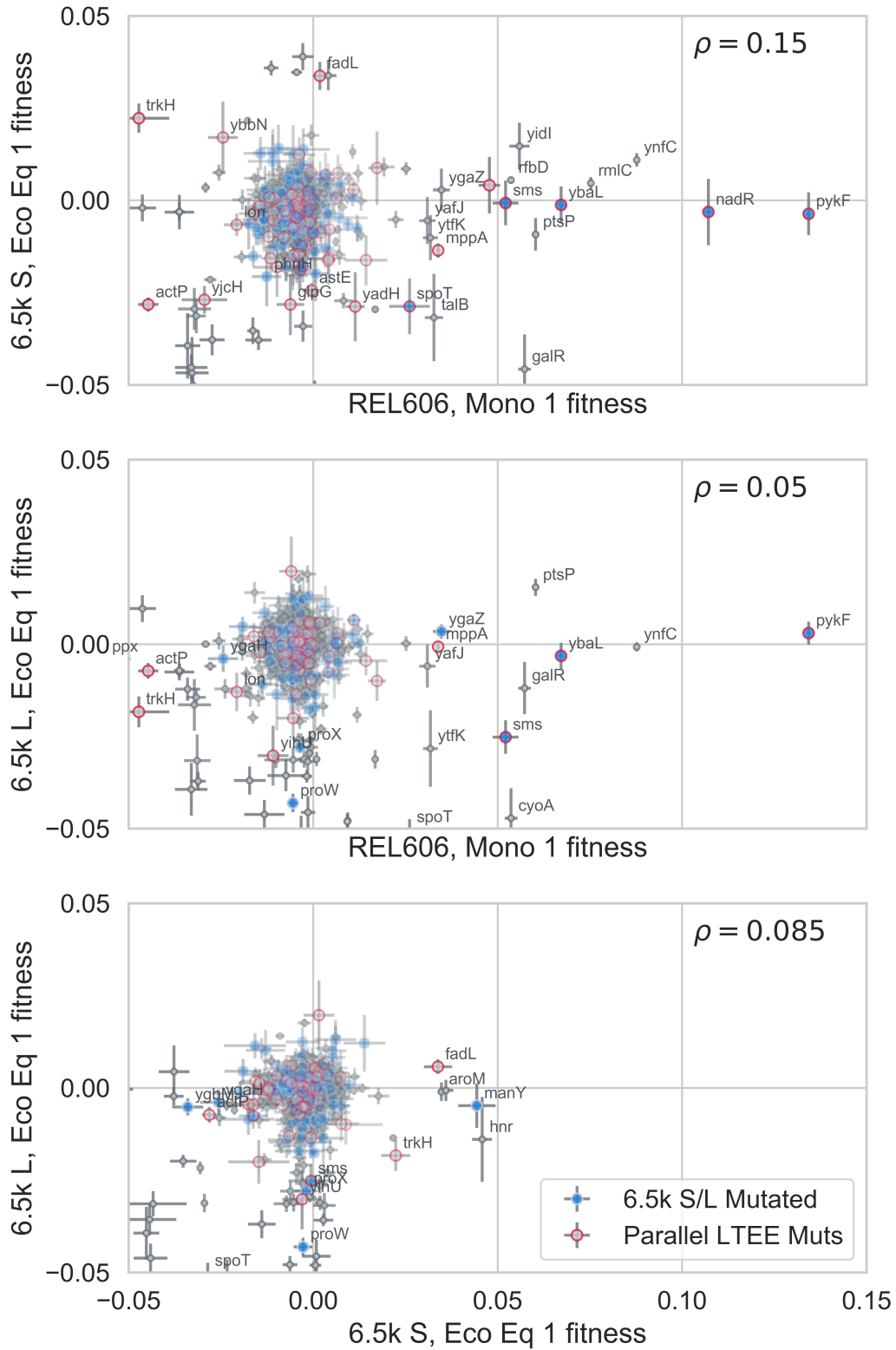


Fig. S9. Comparison of fitness effects; identical to Figure 1F in the main text, except we highlighted all genes in mutated operons. It is still the case that there are many genes that did not get a mutation in their operon, but still changed from a beneficial to non-beneficial fitness effect across genetic backgrounds.

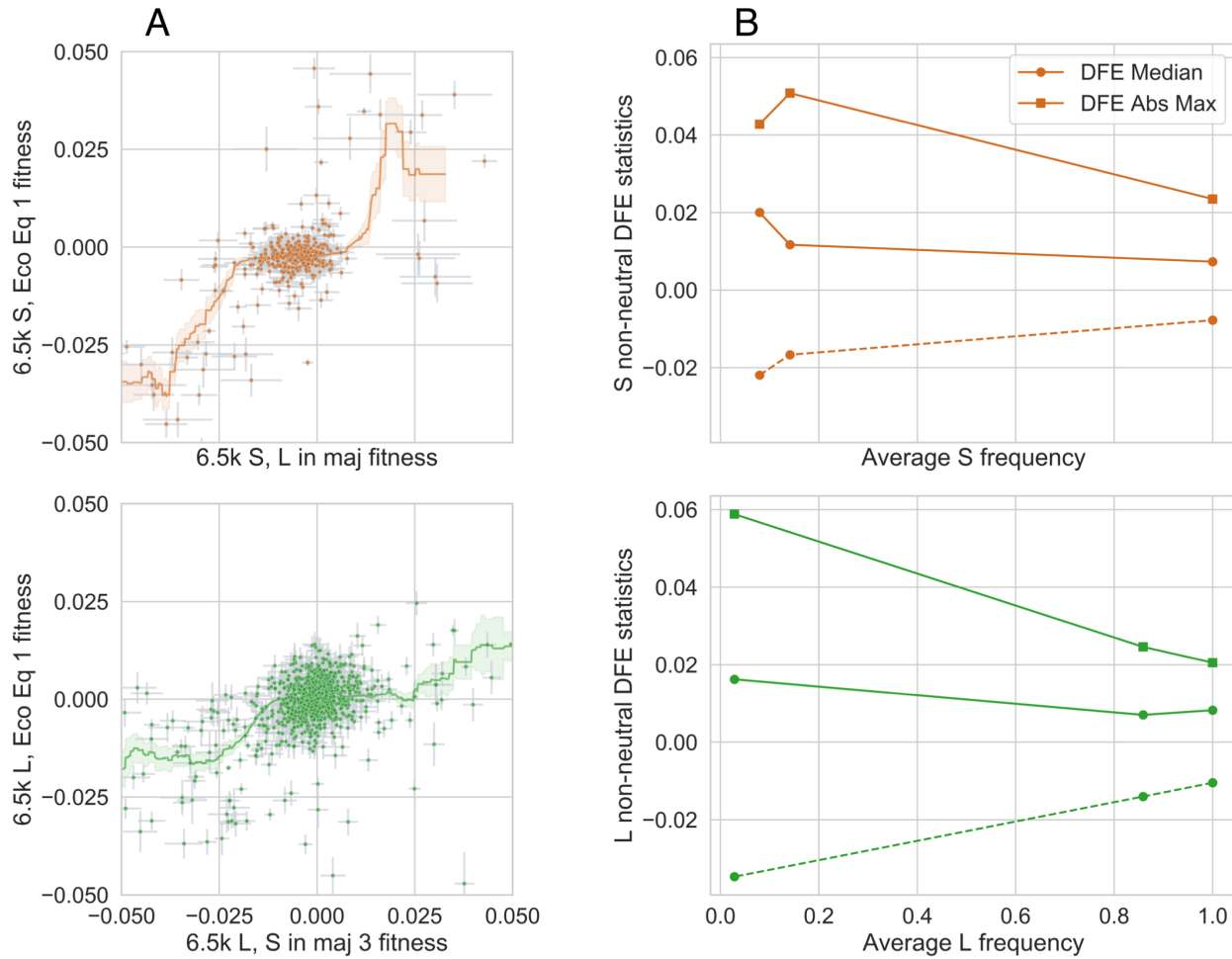


Fig. S10. Frequency-dependent knockout fitness effects for both 6.5k S and L. **(A)** Similar to Fig 2A in main text, except comparing fitness at ecological equilibrium to fitness when the ecotype is in the minority. **(B)** Changes in summary statistics of the DFE as a function of ecotype frequency. Solid lines represent the beneficial side of the DFE, while dashed lines represent the deleterious side.

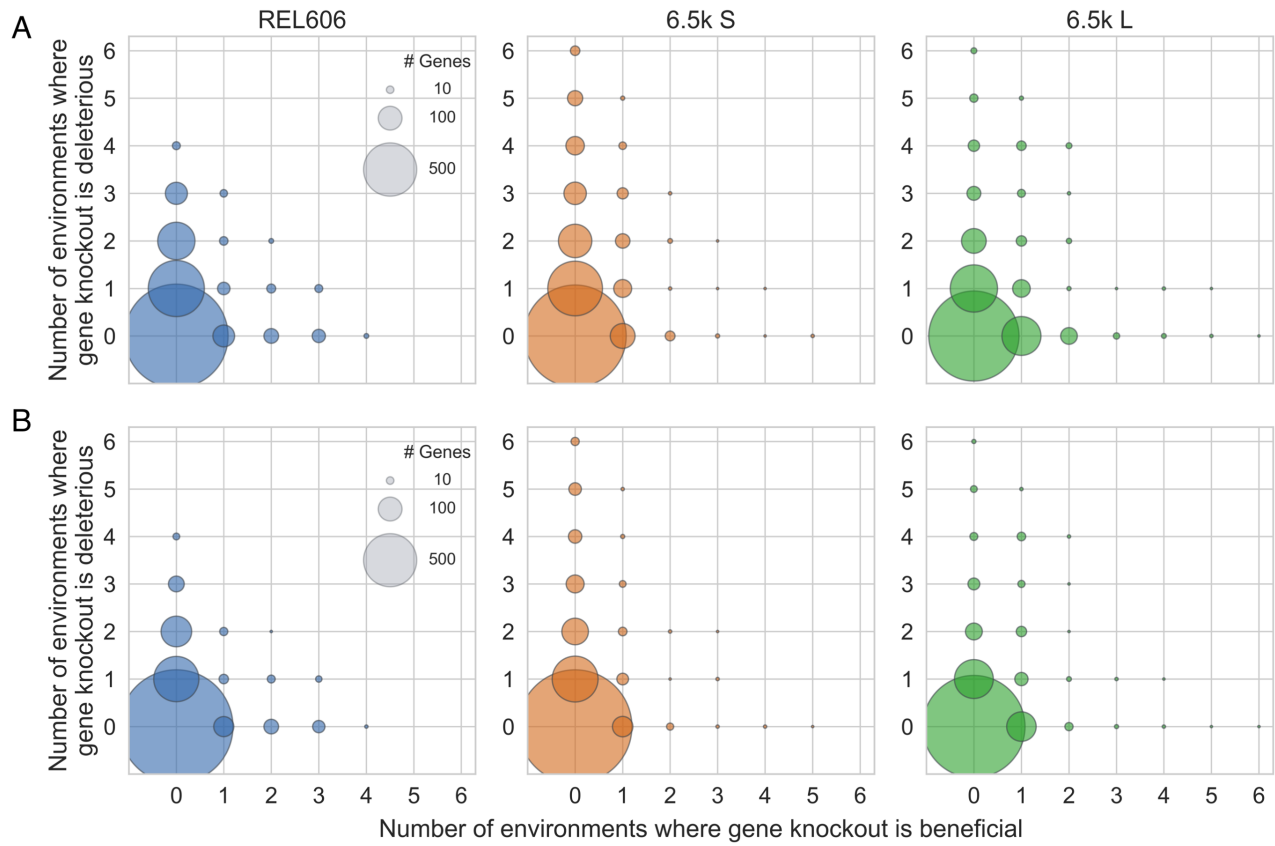


Fig. S11. Fitness effect sign-flipping across environments. Same as Figure 2C (main text), but (post-FDR correction) p-value cutoff is reduced from 0.05 to (A) 10^{-3} or (B) 10^{-5}

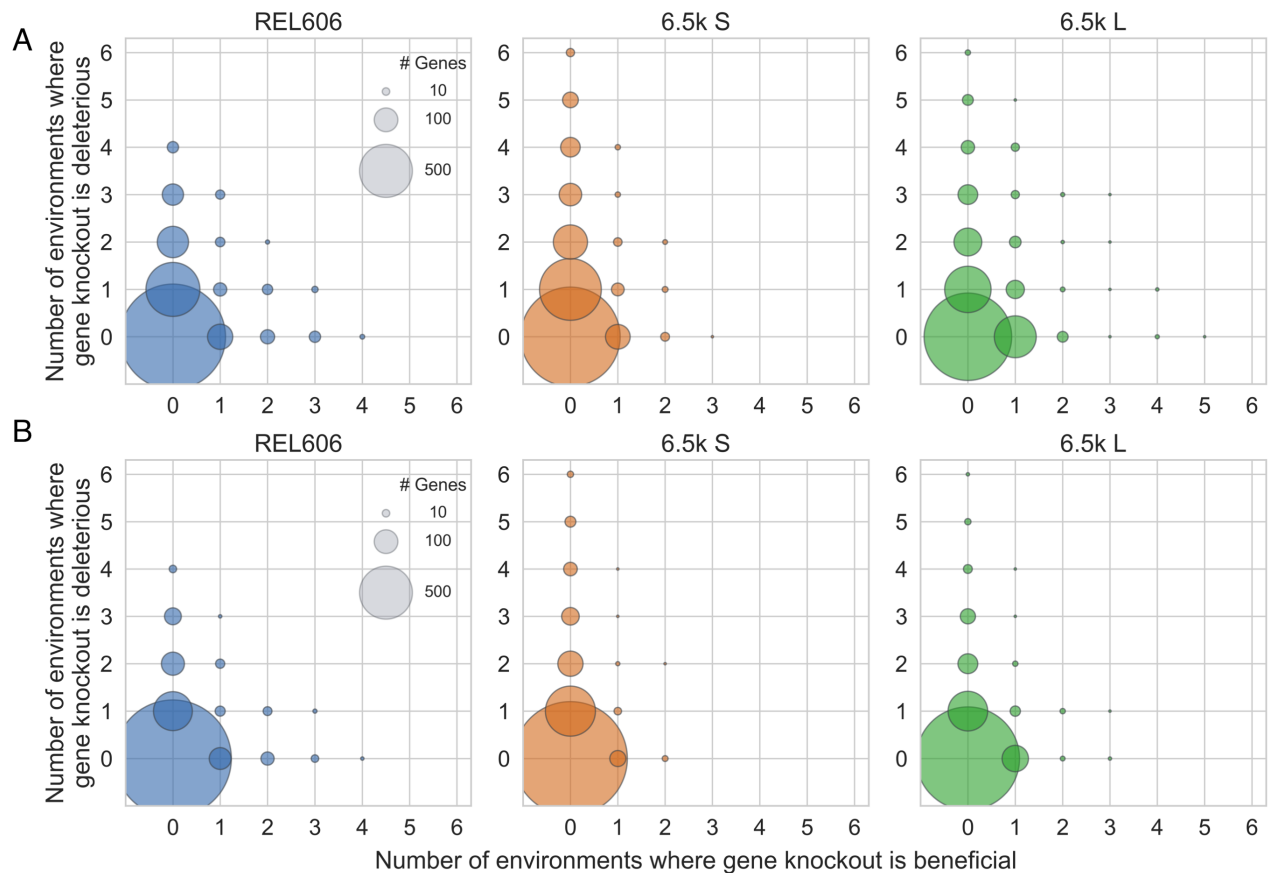


Fig. S12. Fitness effect sign-flipping across environments. Same as Figure 2C (main text), but we only consider genes non-neutral with fitness (A) $|s| > 1\%$ or (B) $|s| > 2\%$

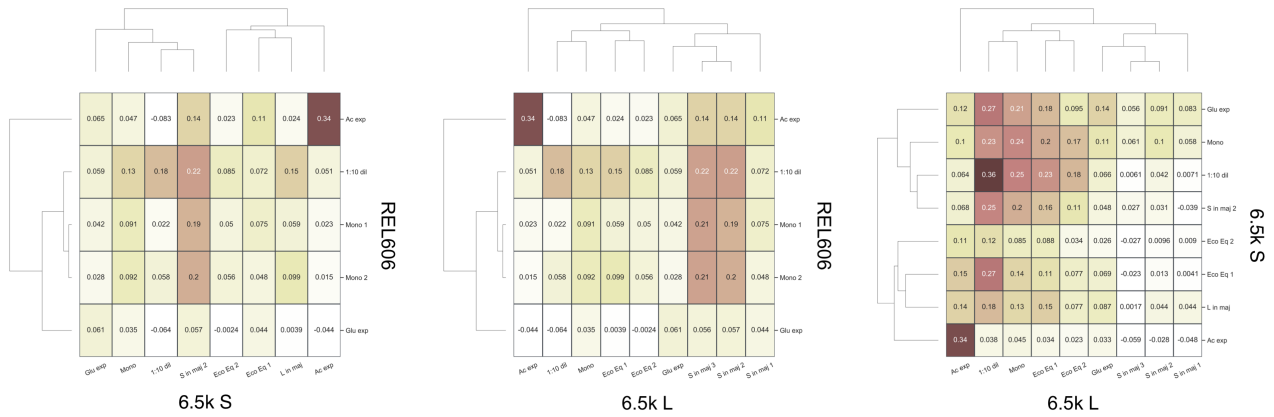


Fig. S13. Fitness effect correlations between strains.

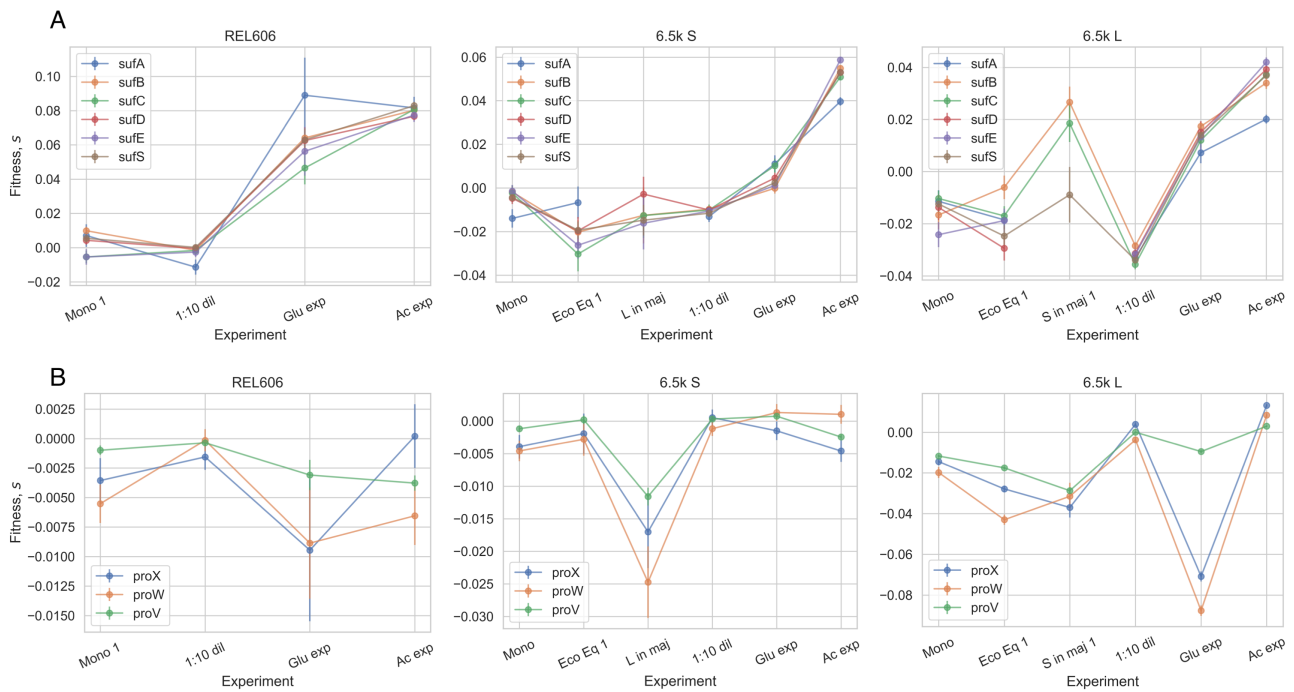


Fig. S14. Fitness effects of (A) *sufABCDSE* and (B) *proVWX* operons in REL606 and 6.5k L/S.

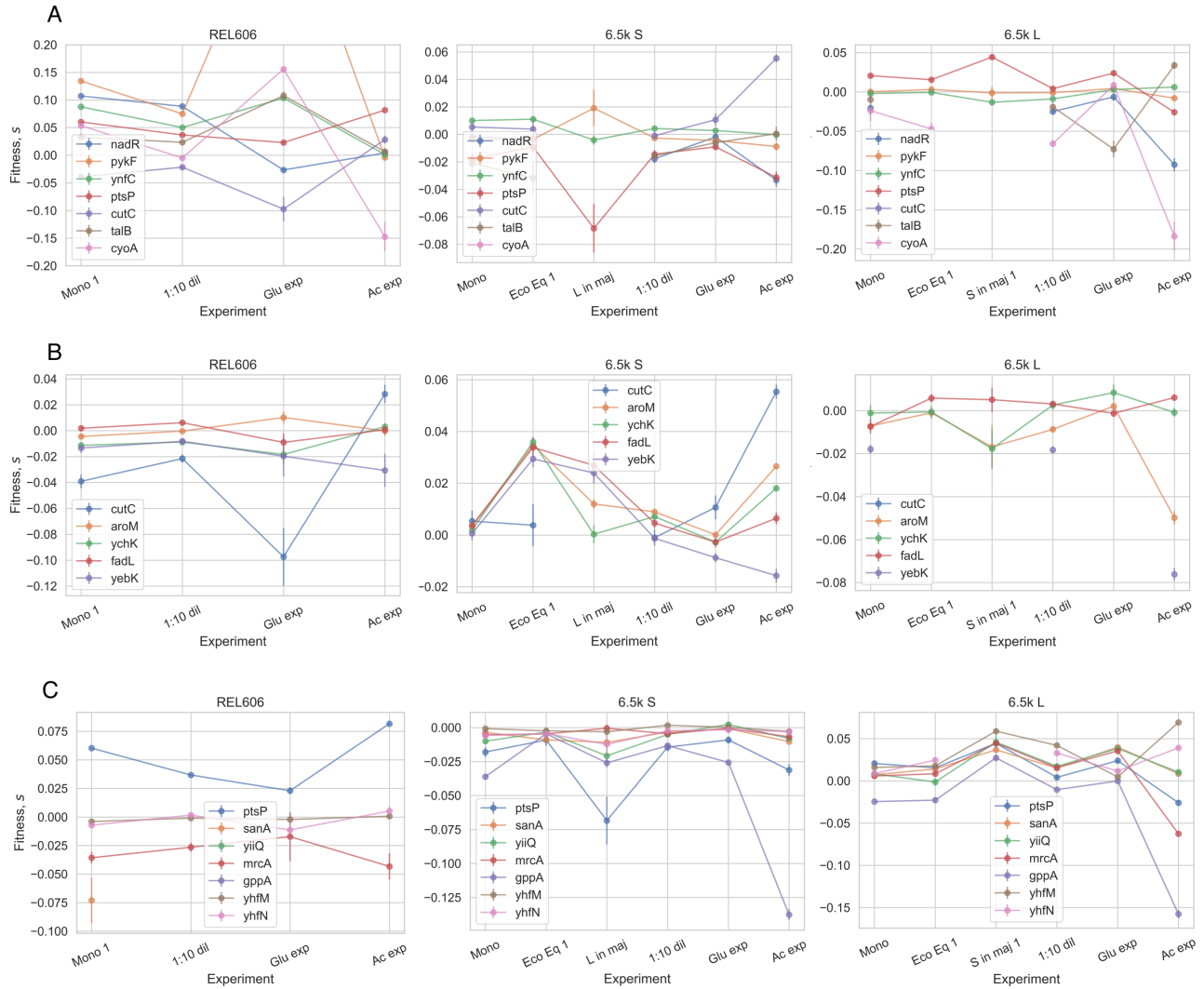


Fig. S15. Fitness effects of knockouts across environments, where knockouts are beneficial in at least one condition on the (A) REL606, (B) 6.5k S, (C) 6.5k L background.

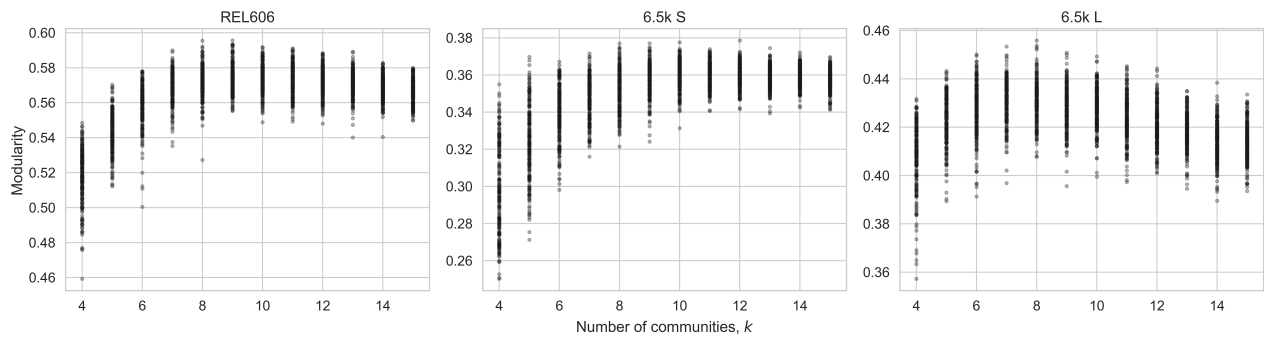


Fig. S16. Modularity of cofitness clusters, across 200 (stochastic) initializations for different numbers of communities from 4 – 15.

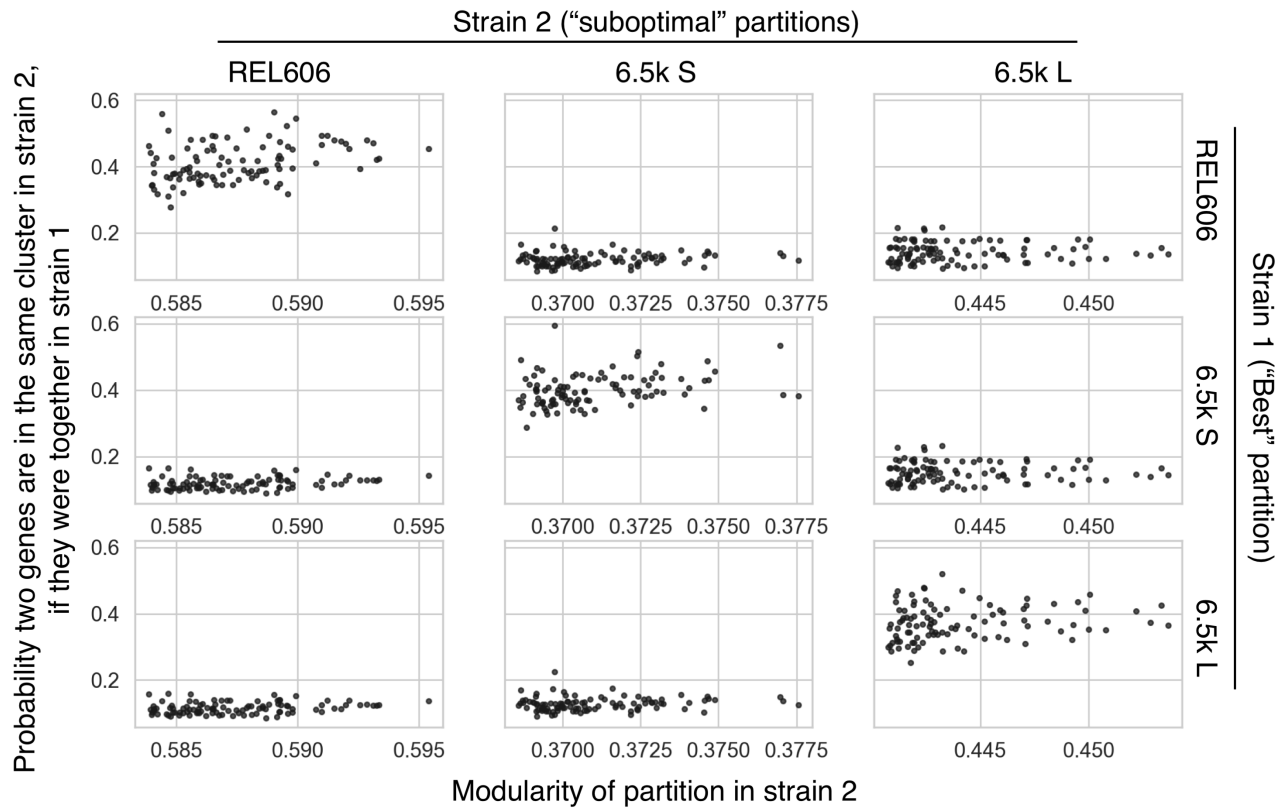


Fig. S17. We compared the optimal partition of the REL606/S/L cofitness networks to the next 100 best (but suboptimal) partitions, also shown in Figure S16. For each suboptimal partition, we asked if two genes were in the same cluster in the optimal partition, what is the probability that they are also in the same cluster in the suboptimal partition. We can see that if we compare partitions in the same genetic background, this probability is around 40%, while it is around 10% when comparing partitions across background. This suggests that different reasonable partitions of the cofitness networks are much more similar within genetic backgrounds than between backgrounds.

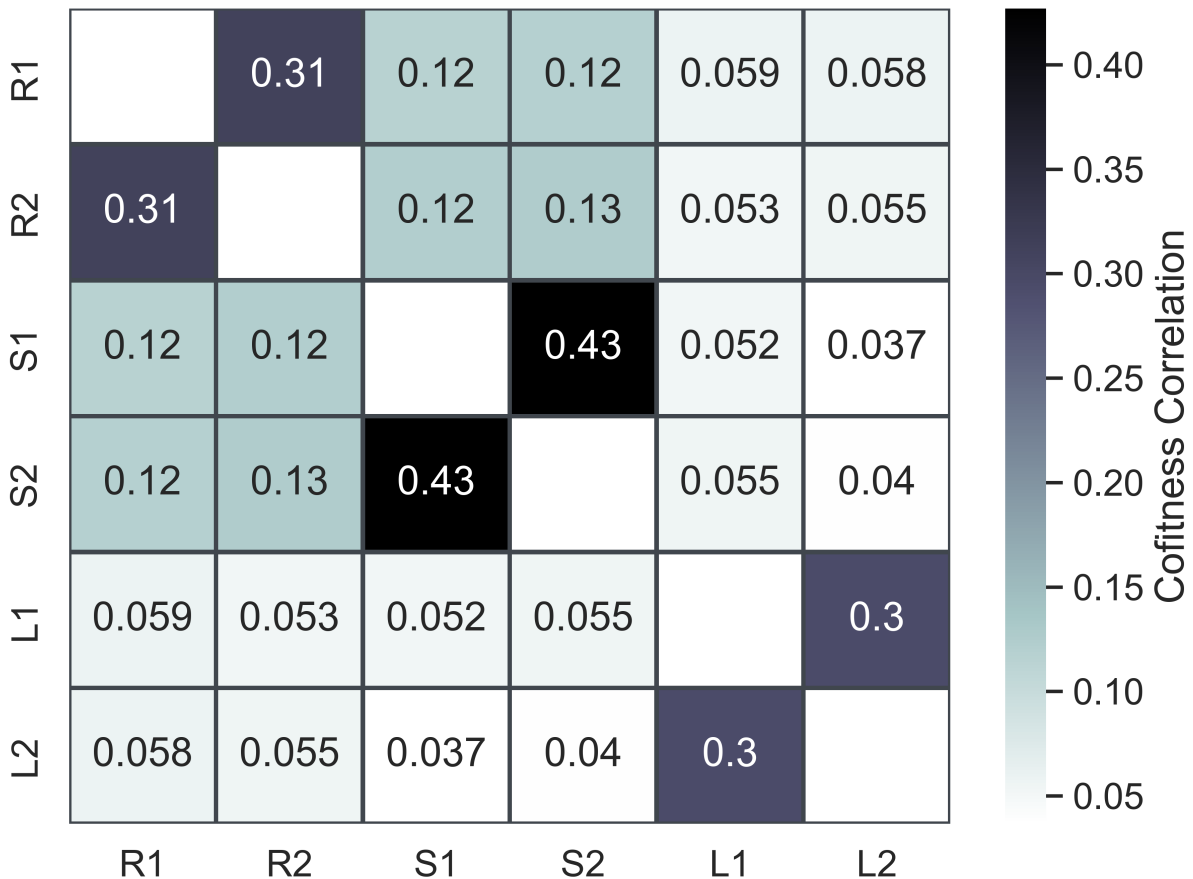


Fig. S18. In order to better understand the extent to which the structure of our cofitness networks is driven by measurement noise, we re-computed the cofitness networks, only using one of the biological replicates per experiment for every experiment. We then computed the correlation of all cofitness values across all networks. We can see that even when the data is independently split, the cofitness networks within a genetic background are more similar than between backgrounds. In the figure, R, S and L refer to REL606, 6.5k S/L libraries, respectively, and 1 and 2 refer to using only biological replicates "1" and "2" from each experiment.

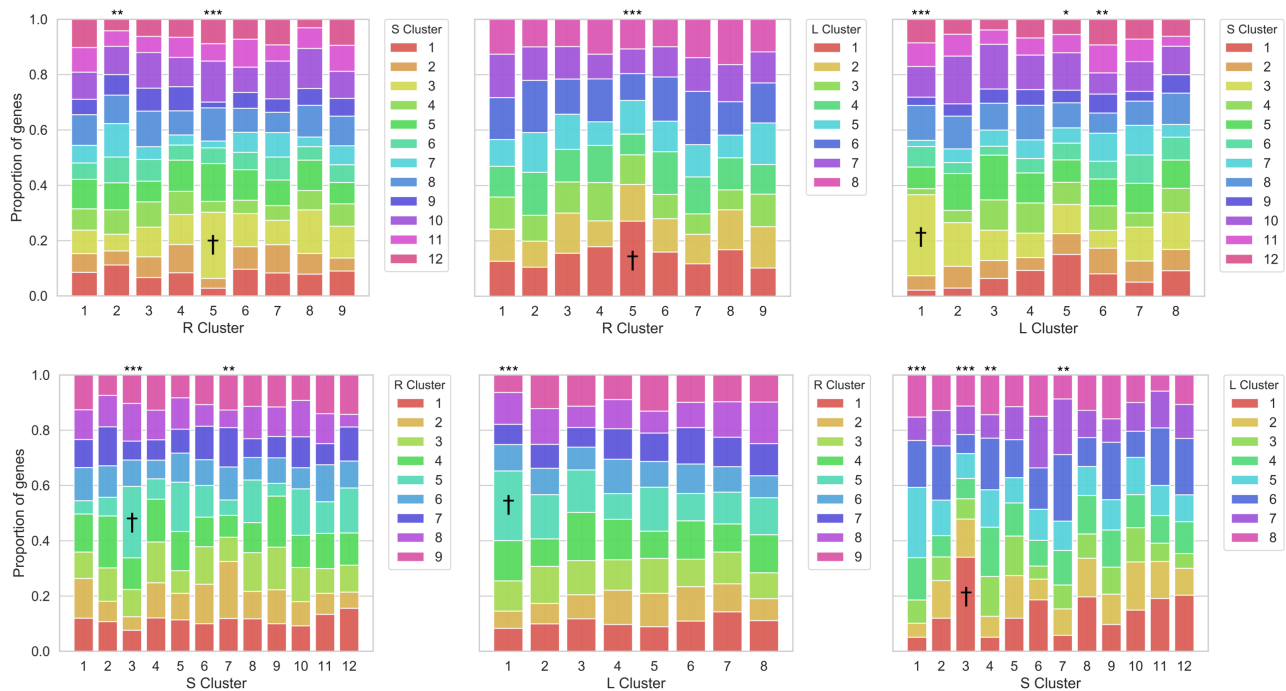


Fig. S19. In order to explore how clusters of genes changed across genetic background, we calculated the fraction of genes in a given cluster that belong to a cluster in a different genetic background. We see that clusters are mostly not preserved between genetic backgrounds, with the exception of the clusters marked by asterisks, which show non-random sampling across genetic backgrounds (* $p < 0.05$, ** $p < 0.005$, *** $p < 10^{-4}$; Pearson's chi-squared test). † set of clusters across all three genetic backgrounds which share more genes than expected, driven primarily by adhesion-related genes.

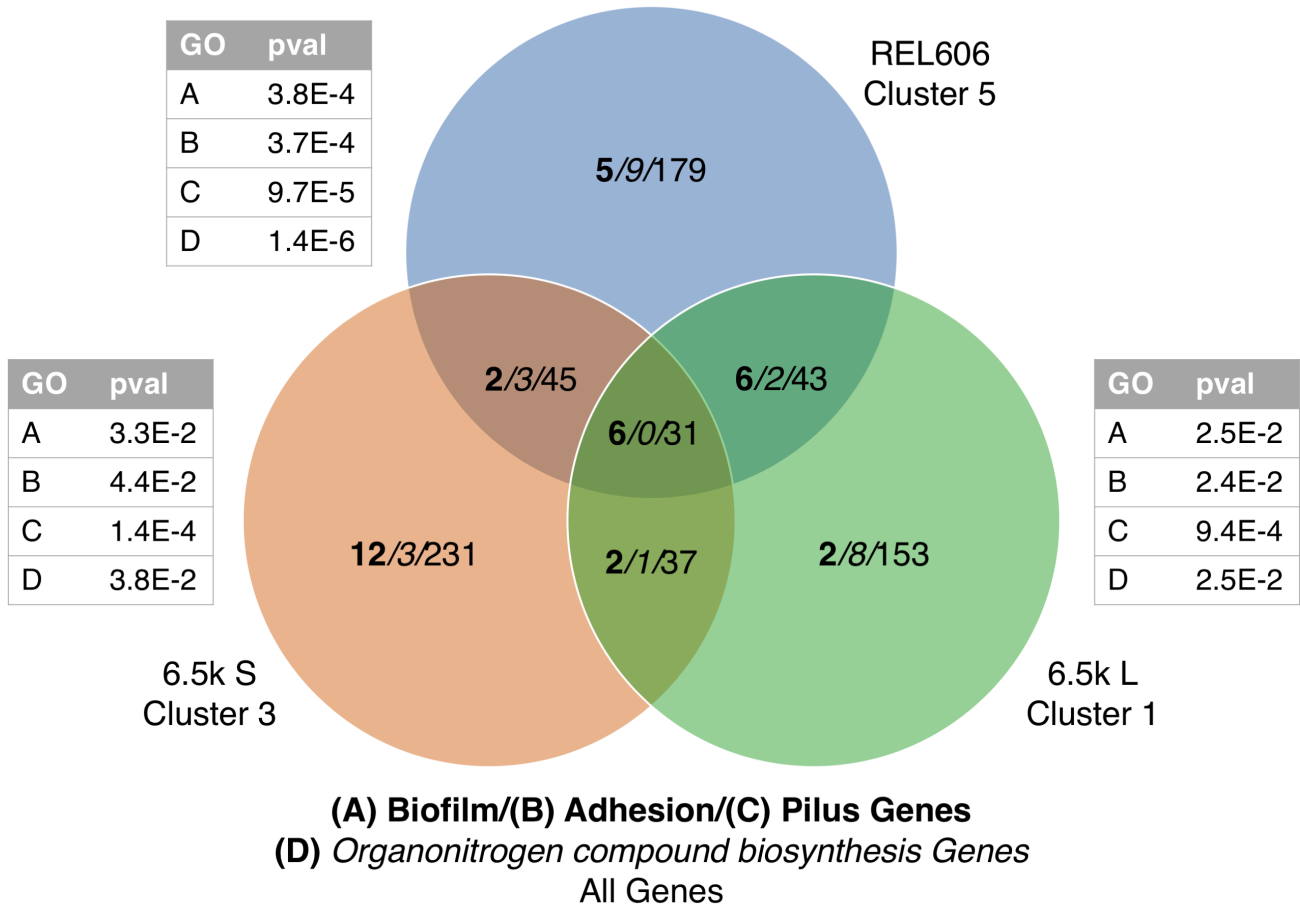


Fig. S20. Biofilm (GO:0043708)/adhesion (GO:0022610)/ pilus organization (GO:0043711)/ organonitrogen compound biosynthesis (GO:1901566) genes tend to appear in the same clusters across genetic backgrounds. P-values are post-FDR correction.

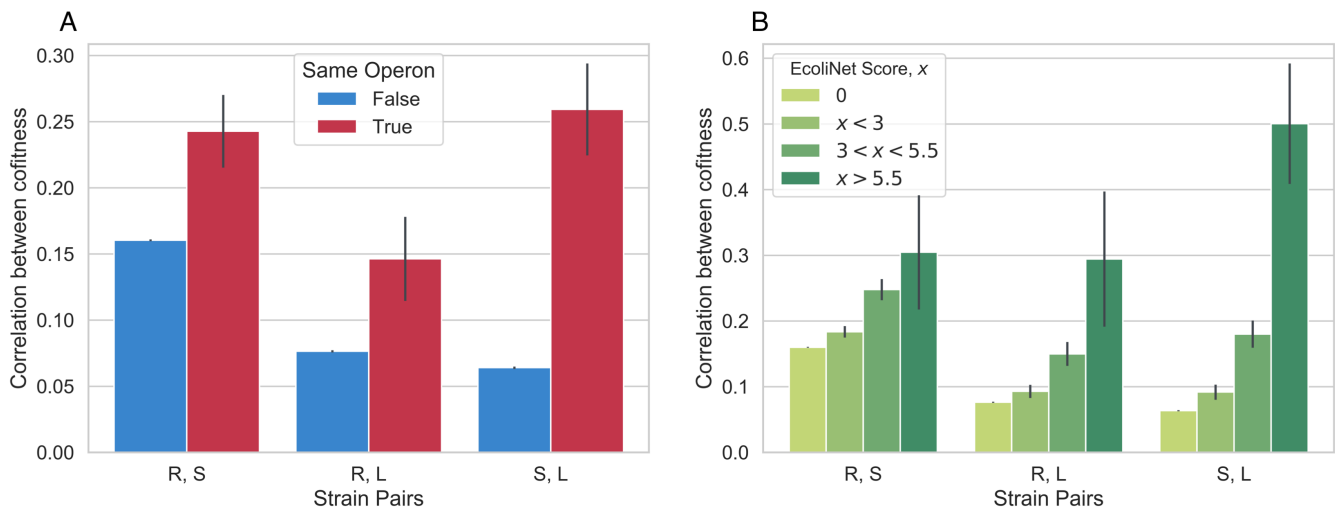


Fig. S21. Correlations between cofitness increase **(A)** when genes are in the same operon and **(B)** with EcoliNet (45) score. A score of 0 indicates that the gene pair is not connected in EcoliNet, i.e. a node distance greater than 1.

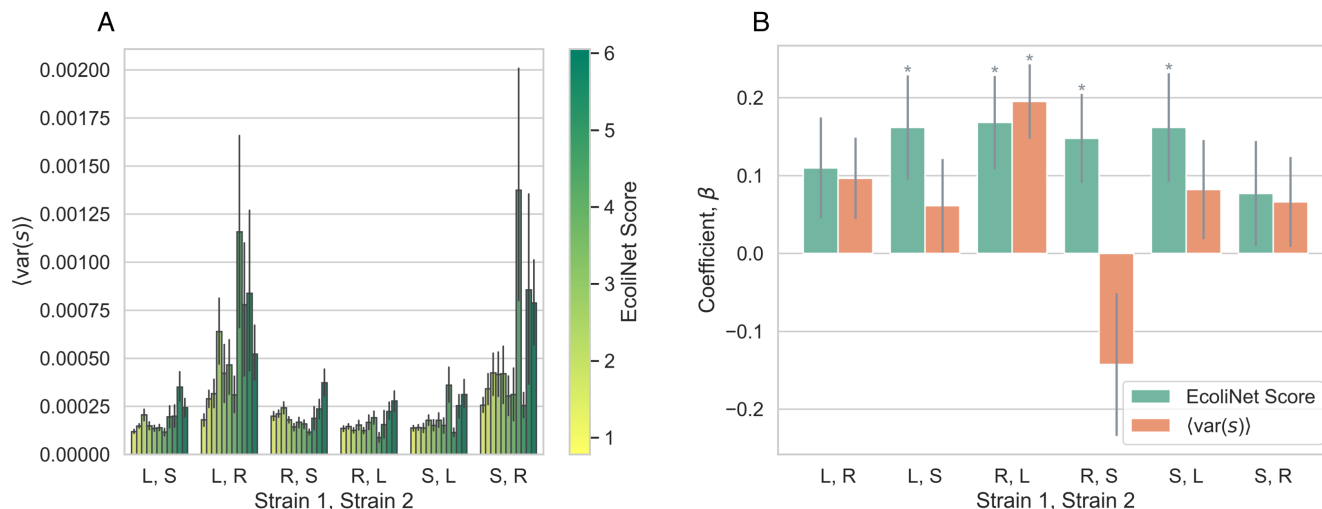


Fig. S22. Variance in fitness effect across environment does not fully explain correlation between EcoliNet score and probability that two genes will be in the same cluster across strains. **(A)** Covariation of EcoliNet scores and variations in fitness effects in some strain pairs. The observed covariation is interesting in and of itself, as it suggests that more strongly interacting genes tend to have a larger variation in fitness effects across environments. **(B)** A standard multiple logistic regression with both fitness variance (in strain 2) and EcoliNet score as covariates, with response variable as the probability two genes are in the same cluster in strain 2, if they were together in strain 1. For most strain pairs, the regression reveals that the correlation reported in Figure 4D still holds after controlling for variation in fitness effects. * $p < 0.05$. See section S4.2 for model details.

Probability two genes are in the same cluster in strain 2, if they were together in strain 1

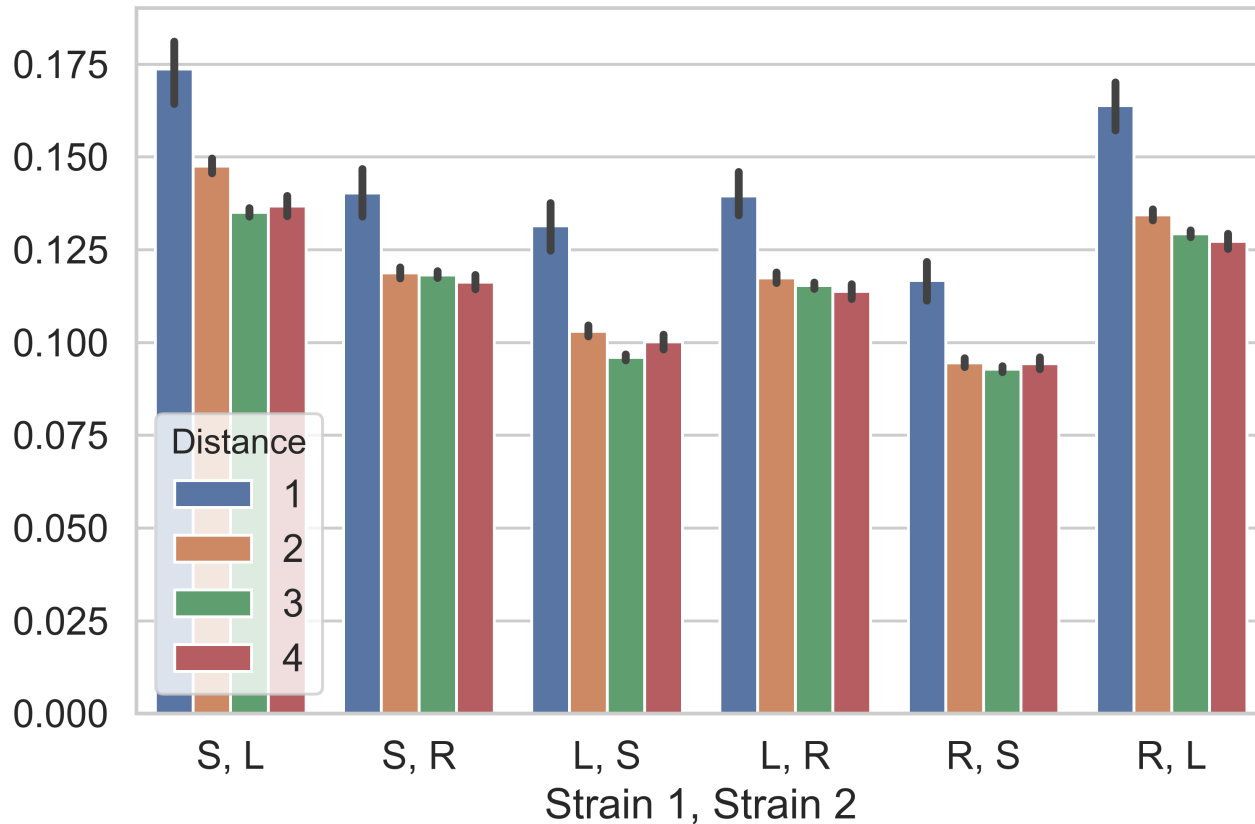


Fig. S23. Shortest distance between genes in EcoliNet (45) predicts if genes stay correlated across genetic background



Fig. S24. Relationship between deleterious mutations and evolutionary outcomes. **(A)** Deleterious fitness effects generally do not predict which genes will mutate, with the one exception that L seems more likely than random to get mutations in genes with deleterious acetate knockout fitness. **(B)** In REL606, deleterious knockout fitness effects are predictive of increased gene expression across all tested environments. In S and L, deleterious fitness effects tend to be associated with an increase in gene expression between 0 and 6.5k generations. Asterisks denote coefficients that are significantly different than 0 (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

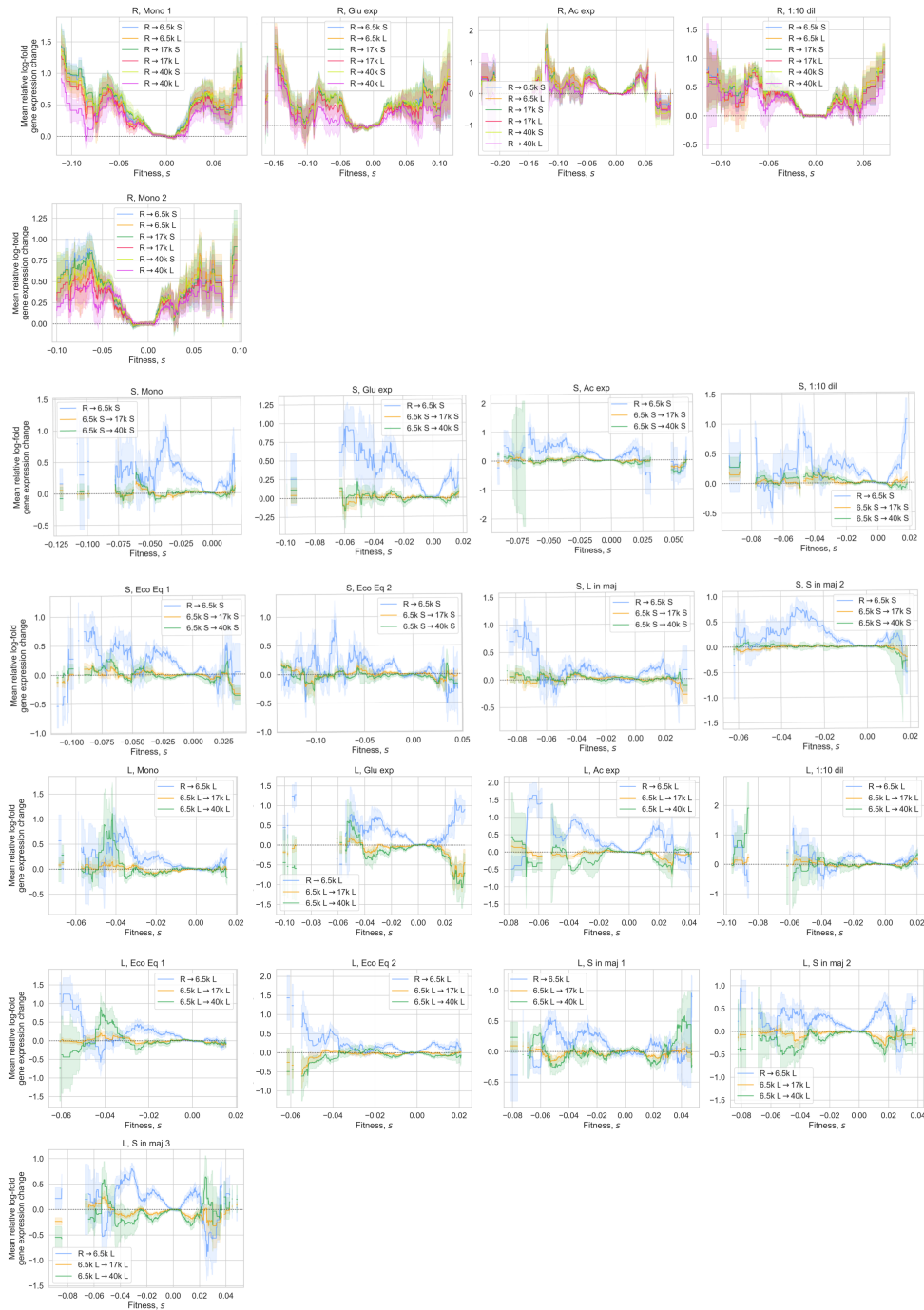


Fig. S25. Relationship between fitness effects and log-fold gene expression change for all experiments, relative to the average change for neutral knockouts. Lines show mean gene expression change as a function of fitness effect (\pm standard error), with a ± 0.01 moving average.

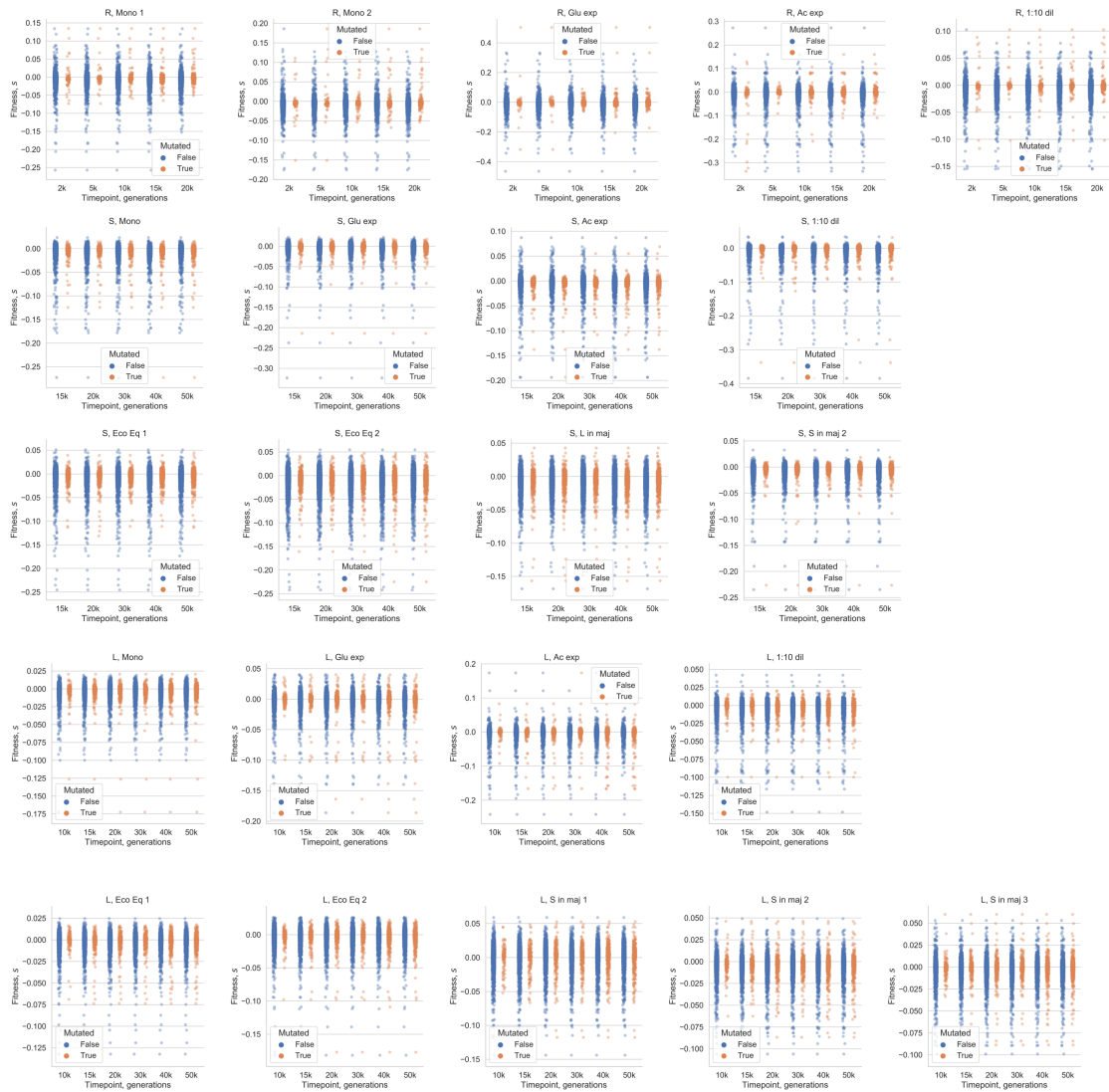


Fig. S26. Establishment of a mutation in a gene by its knockout fitness.

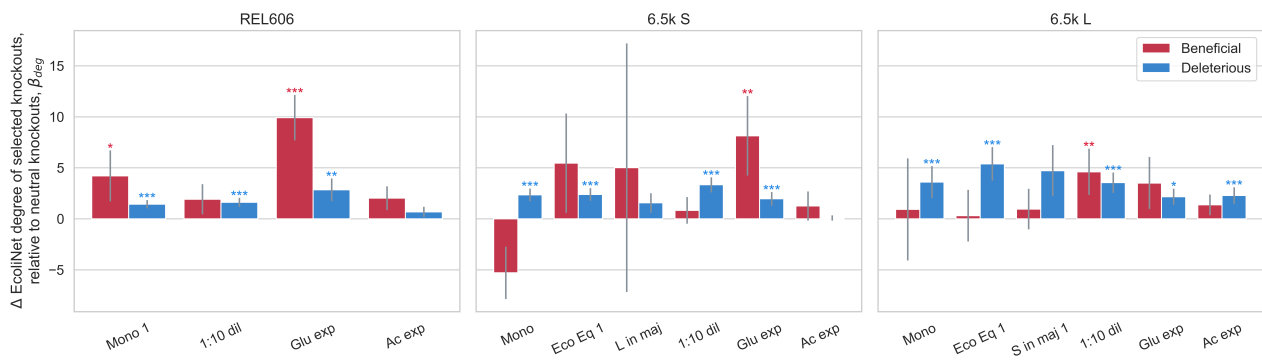


Fig. S27. Fitness effects predict EcoliNet node degree. Deleterious knockouts across environments are more likely to have a high degree compared to neutral knockouts. The same general pattern appears for beneficial knockouts, although less clearly. Linear model fit with ordinary least squares; normalized analogous to model in section S4.4. Asterisks denote coefficients that are significantly different than 0 (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).