

# 1 Hitchhiking and epistasis give rise to cohort dynamics in adapting populations

2

3 Sean W. Buskirk<sup>1</sup>, Ryan Emily Peace<sup>2</sup>, and Gregory I. Lang<sup>1</sup>

4

5 <sup>1</sup>*Department of Biological Sciences, Lehigh University, Bethlehem PA 18015*

6 <sup>2</sup>*Program of Bioengineering, Lehigh University, Bethlehem PA 18015*

7

8

9

10 Keywords: Experimental Evolution, Cohorts, Fitness, Epistasis

11

12

13

## 14 **ABSTRACT**

15 **Beneficial mutations are the driving force of adaptive evolution. In asexual populations, the**  
16 **identification of beneficial alleles is confounded by the presence of genetically-linked hitchhiker**  
17 **mutations. Parallel evolution experiments enable the recognition of common targets of**  
18 **selection, yet these targets are inherently enriched for genes of large target size and mutations**  
19 **of large effect. A comprehensive study of individual mutations is necessary to create a realistic**  
20 **picture of the evolutionarily significant spectrum of beneficial mutations. Here we utilize a bulk-**  
21 **segregant approach to identify the beneficial mutations across 11 lineages of experimentally-**  
22 **evolved yeast populations. We report that most genome sequence evolution is non-adaptive:**  
23 **nearly 80% of detected mutations have no discernable effects on fitness and less than 1% are**  
24 **deleterious. We determine the distribution of driver and hitchhiker mutations in 31 mutational**  
25 **cohorts, groups of up to ten mutations that arise synchronously from low frequency and track**  
26 **tightly with one another. Surprisingly, we find that one-third of cohorts lack identifiable driver**  
27 **mutations. In addition, we identify intra-cohort synergistic epistasis between mutations in *hs17***  
28 **and *kel1*, which arose together in a low frequency lineage.**

29

## 30 INTRODUCTION

31 Adaptation is a fundamental biological process. The identification and characterization of the  
32 genetic mechanisms underlying adaptive evolution remains a central challenge in biology. To identify  
33 beneficial mutations, recent studies have characterized thousands of first-step mutations and systemic  
34 deletion and amplification mutations in the yeast genome (1, 2). These unbiased screens provide a  
35 wealth of information regarding the spectrum of beneficial mutations, their fitness effects, and the  
36 biological processes under selection. However, this information alone cannot predict which mutations  
37 will ultimately succeed in an evolutionary context as genetic interactions and population dynamics also  
38 impart substantial influence on the adaptive outcomes.

39 Early theoretical models assume that beneficial mutations are rare, such that once a beneficial  
40 mutation escapes drift, it will fix (3-5). For most microbial populations, however, multiple beneficial  
41 mutations will arise and spread simultaneously, leading to complex dynamics of clonal interference and  
42 genetic hitchhiking (6-9), and in many cases, multiple mutations track tightly with one another through  
43 time as mutational cohorts (10-13). The fate of each mutation is therefore dependent not only on its  
44 own fitness effect, but on the fitness effects of and interactions between all mutations in the population.  
45 Many beneficial mutations will be lost due to drift and clonal interference while many neutral (and even  
46 deleterious) mutations will fix by hitchhiking. The influence of clonal interference and genetic hitchhiking  
47 on the success of mutations makes it difficult to identify beneficial mutations from sequenced clones or  
48 population samples. Identifying beneficial mutations and quantifying their fitness effects ultimately  
49 requires assaying each individual mutation independent of co-evolved mutations.

50 Here we present a comprehensive, large-scale survey that quantifies the fitness effects of 116  
51 mutations from eleven evolved lineages for which high-resolution knowledge of the dynamics of  
52 genome sequence evolution is known. We describe a novel strategy for constructing bulk-segregant  
53 pools that enables high-throughput quantification of fitness effects of individual evolved alleles. We find  
54 that large-effect mutations in common targets of selection drive adaptation while deleterious alleles  
55 rarely reach appreciable frequencies. We report that most genome sequence evolution is non-adaptive:  
56 nearly 80% of detected mutations have no discernable effect on fitness. Furthermore, one-third of  
57 cohorts lack identifiable driver mutations, and the dynamics of these “driverless” cohorts can be  
58 explained by genetic hitchhiking alone. Through the extensive characterization of evolved mutations,  
59 we begin to explore the mechanisms responsible for the observed cohort dynamics and we identify one  
60 cohort that suggests that rare mutations and epistatic interactions represent evolutionarily-relevant  
61 genetic mechanisms of adaptation.

62

## 63 RESULTS

### 64 Fitness of eleven representative lineages from nine evolved populations

65 Previously we evolved nearly 600 haploid *MATa* populations for 1,000 generations in a rich glucose  
66 medium. These populations were divided between two strains and two population sizes. “RM” and “BY”  
67 strains differ by a single SNP in *GPA1* that modifies the benefit of mutations in the mating pathway—a  
68 trait that we have studied extensively previously (8, 14). Different population sizes were maintained by  
69 controlling the dilution frequency and bottleneck size. We selected clones from both RM and BY  
70 populations for which we previously followed the frequencies of all mutations at high temporal  
71 resolution, selecting only those evolved at the smaller population size ( $N_e \sim 10^5$ ) to maintain uniformity  
72 across our experiments in this study. Since the number and identity of mutations in each population is  
73 known, we sampled across the range of biological pathways under selection (i.e. mating pathway, Ras  
74 pathway, cell wall assembly/biogenesis) and captured clones that exhibit a range of genome  
75 divergence, from 3 mutations in BYS1A08 to 16 mutations in BYS1E03. In two instances (RMS1G02  
76 and RMS1H08), clones were isolated from two co-existing subpopulations to investigate how particular  
77 mutations and their fitness effects impact the ultimate fate of competing lineages. We identify selected  
78 clones by their population and the timepoint from which they were isolated (for example, clone  
79 BYS1A08-545). The evolutionary dynamics corresponding to each of the eleven selected clones are  
80 shown in their entirety in Figs. S2 and S3. We measured the fitness of each clone against a  
81 fluorescently-labeled ancestral strain of the appropriate background (BY or RM). Fitness values ranged  
82 from 2.7% in RMS1G02-545 to 8.6% in RMS1D12-910 (Fig. S1 and Table S1).

### 83 Novel strategy for constructing bulk-segregant pools

84 We developed a novel bulk-segregant approach to rapidly and efficiently generate large pools of  
85 *MATa* segregants that contain random combinations of evolved mutations (Fig. S2). Briefly, we  
86 backcrossed evolved clones to a *MAT $\alpha$* -version of the ancestor, gene converted the mating-type locus,  
87 and sporulated the resulting *MATa/a* diploid by complementing the *MAT $\alpha$ 2* gene on a plasmid. Our  
88 method has three key advantages over the commonly used yeast “magic marker” approach, which  
89 selects on auxotrophic markers driven by mating-type specific promoters (15). First, our strategy  
90 ensures that the segregant pool is strictly *MATa* through the removal of the *MAT $\alpha$*  gene during pool  
91 construction. Second, the approach produces segregants that are nearly isogenic to the evolved strain,  
92 thereby avoiding undesirable genetic interactions with the magic-marker machinery. Finally, the method  
93 can be applied to any strain background without requiring the incorporation of the magic-marker  
94 machinery and coordination of auxotrophic reporters.

## 95 **Fitness effect of 116 mutations across 11 evolved lineages**

96 For each of our eleven evolved clones we generated a pool of  $\sim 10^5$  haploid *MATa* segregants. We  
97 propagated each pool for 90 generations and determined the fitness of each mutation by sequencing to  
98 a depth of  $\sim 100$  reads per site every 20 generations (Fig. S2). The background-averaged fitness is  
99 calculated as the slope of the  $\ln(\text{mutant allele count/wild-type allele count})$  versus generations. In  
100 addition, we determined the segregation of fitness within each of the eleven crosses by isolating 192  
101 segregants from each pool and quantifying fitness using a flow cytometry-based competitive fitness  
102 assay for a total of 2,112 fitness assays on individual segregants.

103 Two representative lineages are shown in Fig. 1, and all eleven lineages are shown in Fig. S1. We  
104 analyzed single clones from populations BYS2D06 and RMS1D12, each isolated from Generation 910.  
105 The evolutionary history of population BYS2D06 shows that eleven mutations swept through the  
106 population as three independent cohorts (Fig. 1) spaced several hundred generations apart. Bulk-  
107 segregant analysis revealed the presence of three beneficial mutations (*gas1*, 3.4%; *ira2*, 2.7%; and  
108 *ste5*, 3.3%) in the clone BYS2D06-910, each driving a single cohort (Fig. 1). The fitness distribution of  
109 the individual segregants shows distinct modes that fall within the range of fitness bounded by the  
110 ancestral and evolved parental strains. Population RMS1D12 exhibits more complicated dynamics (Fig.  
111 1). The clone RMS1D12-910 contains fourteen mutations grouped into six cohorts (Fig. S3). Bulk-  
112 segregant analysis identifies three beneficial mutations (*ira1*, 3.3%, *rot2*, 5.6%, and *mid2*, 2.1%), each  
113 driving a different cohort. Again, the fitness values of the individual segregants fall between the fitness  
114 of the two parental strains, though the distribution appears more bimodal and less continuous than the  
115 distribution of fitness among the BYS2D06-910 segregants. These two major modes likely represent  
116 those segregants with and without the 5.6% fitness-effect *rot2* mutation.

117 Across the eleven lineages we measured the fitness of 116 of the 120 expected mutations. For four  
118 of the mutations we were unable to quantify fitness because all of the segregants in the pool contain  
119 the same allele due to gene conversion events during strain construction (see Discussion below). Of  
120 the 116 mutations across these eleven lineages, we find that 24 of the 116 mutations are beneficial,  
121 ranging in fitness from  $\sim 1\%$  to  $\sim 10\%$  effect. This indicates that most genome evolution is non-adaptive:  
122 roughly 80% of detected mutations are hitchhikers. As expected, all evolved clones possess at least  
123 one beneficial mutation. In addition, evolved clones contain up to 13 hitchhiker mutations (Fig. 2 and  
124 Table S2). Three mutations are reported as deleterious, though only one, a read-through mutation in  
125 the stop codon of *gcn2* in population BYS1D08, has greater than a 1% effect on fitness. We find no  
126 evidence of synonymous or intergenic mutations increasing fitness. When parsed by mutation type,  
127 driver mutations are enriched for nonsense and frameshift mutations compared to hitchhiker mutations

128 (Fig. 3A). In contrast, synonymous and intergenic mutations are exclusively hitchhikers. This is  
129 consistent with adaptation driven by loss-of-function mutations in coding sequences (16).

### 130 **Mapping fitness effects to the dynamics of adaptation**

131 Mutations often move through the populations as cohorts, synchronously escaping drift and tracking  
132 tightly with one another through time (10). Cohorts are a recent observation (10-13) and the  
133 evolutionary dynamics that drive their formation are yet to be explained. In general, mutations within a  
134 cohort are genetically and functionally unrelated, suggesting that cohorts are a random collection of  
135 mutations that accumulate while at low frequencies on a common background. A deeper understanding  
136 of these dynamics requires resolving the number of driver and hitchhiker mutations for a large sampling  
137 of cohorts. We used hierarchical clustering to objectively partition the 116 mutations into distinct cohorts  
138 based on correlated changes in allele frequency over time. Across the eleven selected lineages, we  
139 identify 31 distinct cohorts (Fig. S3). Each cohort contains up to ten mutations and each evolved  
140 lineage possesses up to six cohorts. Eighteen cohorts have a single driver mutation. Three cohorts  
141 each possess two unique driver mutations. These two-driver cohorts have twice as many hitchhiker  
142 mutations compared to single-driver cohorts ( $6.0 \pm 2.0$  and  $3.1 \pm 1.9$ , respectively,  $P=0.02$ , Wilcoxon  
143 Rank Sum, one-tailed), contributing to the overall positive correlation between number of beneficial  
144 mutations per cohort and cohort size (Fig. 3,  $\rho=0.70$ , Pearson correlation). Surprisingly, no beneficial  
145 mutations were detected within 10 of the 31 cohorts. These “driverless” cohorts typically contain few  
146 mutations; half are just a single mutation and only one is comprised of more than three mutations (Fig.  
147 3B). We propose that driverless cohorts represent a form of genetic hitchhiking, where non-adaptive  
148 cohorts are “pulled” to intermediate frequency by the preceding cohort and/or “pushed” to higher  
149 frequencies during the sweep of the following cohort (Fig. S4).

### 150 **Synergistic epistasis between *hsl7* and *kel1*, two genes whose protein products localize to sites** 151 **of polarized growth.**

152 In the absence of epistasis, the fitness of an evolved clone is equivalent to the sum of the individual  
153 fitness effects of all constituent mutations. Indeed, we find a positive correlation ( $\rho=0.82$ , Pearson  
154 correlation) between the additive expectation of all mutations affecting fitness and the measured fitness  
155 of each evolved clone (Fig. 3C). Deviations from this additive expectation could indicate epistatic  
156 interactions between mutations that arose in the evolved lineage. One clone in particular, BYS2E01-  
157 745, deviates strongly from this additive expectation. Our bulk-segregant fitness assay identified two  
158 driver mutations in this lineage, *hsl7* (9.2%) and *kel1* (7.1%). The additive expectation (16.3%) is over  
159 twice the experimentally measured fitness of 7.6% (Fig. 3C). In addition, the BYS2E01-745 cross

160 exhibits a highly asymmetric fitness distribution with a large proportion of low fitness segregants. (Fig.  
161 S1).

162 The dynamics of the BYS2E01-745 lineage are simple: an abrupt sweep of a single cohort of  
163 eleven mutations (Fig. 4A). However, the genetic basis of adaptation is unclear since BYS2E01-745 is  
164 the only lineage in this study without a mutation in a putative target of selection (10). Interestingly, the  
165 BYS2E01-745 lineage is enriched for mutations in genes whose products localize to the cellular bud  
166 and site of polarized growth ( $P=0.0012$  and  $P=0.0016$ , respectively; GO Term Finder (17)). Further, all  
167 four genes (*hsl7*, *kel1*, *iqg1*, and *ccw12*), whose protein products localize to sites of polarized growth,  
168 contain either missense or frameshift mutations (Fig. 4A). Taken together, these data suggest that  
169 mutations in the BYS2E01-745 lineage may interact epistatically.

170 To test for epistatic interactions in the BYS2E01-745 lineage we genotyped all 192 segregants  
171 using low-coverage whole-genome sequencing (Table S4). Of the eleven mutations in the BYS2E01-  
172 745 lineage, only two loci significantly affect fitness: *HSL7* and *KEL1* ( $P < 0.0001$ ; N-way ANOVA, Type  
173 III SS), corroborating the results of the bulk-segregant fitness assay (Fig. S1). Furthermore, the  
174 pairwise interaction between the *hsl7* and *kel1* alleles significantly impact fitness ( $P < 0.0001$ ; N-Way  
175 ANOVA, Type III SS). The *hsl7* mutation confers a modest benefit in the *KEL1* wild-type background  
176 (~2.1%), while it confers a significantly larger advantage (~7.9%) when paired with the *kel1* allele (Fig.  
177 4B). The *kel1* mutation is nearly neutral (-0.1%) in the *HSL7* wild-type background, but provides a  
178 substantial benefit in the *hsl7* background (~5.6%). As initially suspected from the bulk-segregant  
179 fitness assay (Fig. S1), the *IQG1* allele is absent from all individual segregants, the result of an  
180 endoduplication event that occurred during the construction of the parental diploid strain (Fig. S5).  
181 Through allelic replacement of *IQG1* in individual segregants, we found that the evolved *iqg1* allele has  
182 no effect on fitness. These results demonstrate that the rise of the BYS2E01-745 lineage is driven by  
183 intra-cohort epistasis between the *hsl7* and *kel1* mutations, which combine to produce a high fitness  
184 genotype, the fittest of the six BY evolved clones (Table S1). Detection and characterization of the *hsl7*-  
185 *kel1* interaction has revealed the only instance of synergistic epistasis to arise within a long-term  
186 evolution experiment outside of the well-studied Cit<sup>++</sup> phenotype in *E. coli* (18-20).

187

## 188 DISCUSSION

189 Recurrence-based models represent a widely-accepted statistical approach for inferring which  
190 genes and biological processes are under selection, by identifying targets that are mutated more often  
191 than expected by chance (21-25). Such probability-based methods, however, will inherently neglect



192 rare driver mutations and are unsuitable for quantifying fitness as occurrence is not necessarily  
193 indicative of fitness effect. Many variables, e.g. mutational target size, can impact how often mutations  
194 in a particular target are detected. Indeed, we find no apparent relationship between the fitness effect  
195 that a driver confers and its prevalence (Fig. S6). For instance, both *ira1* and *ira2* provide similar effects  
196 on fitness (~2.7% each); however, *ira1* mutations were observed 23 times across replicate populations,  
197 yet *ira2* only once. Of our 24 driver mutations, our recurrence model did not detect six because they  
198 were mutated only twice; these include four modest-effect mutations (between 1 and 2%) as well as the  
199 two epistatic interactors, *hsl7* and *kel1*, in population BYS2E01 (Fig. S7).

200 The power behind identification of drivers through the bulk-segregant approach lies in the ability to  
201 screen lineages with numerous divergent loci for adaptive alleles. In only 1,000 generations, our  
202 populations fixed up to 19 mutations and in many cases, multiple cohorts of mutations existed  
203 simultaneously in the population leading to complex evolutionary dynamics. In addition, factors such as  
204 mutational target size, epistasis, genetic hitchhiking, and clonal-interference strongly influence the  
205 identity of mutations that arise over the course of evolution. Therefore, directly measuring the fitness  
206 effects of all mutations, as we do here, is necessary to unambiguously identify and quantify the fitness  
207 effects of driver mutations that could otherwise be missed by recurrence-based methods.

208 There is increasing evidence that large-effect beneficial mutations drive adaptation in microbial  
209 evolution experiments (1, 2) as well as in natural microbial populations (26) and clinical infections (27).  
210 Here we identify 24 driver mutations ranging from roughly 1% to 10% effect, consistent with populations  
211 evolving under the strong-selection strong-mutation paradigm in which small fitness effects (<1%) are  
212 unlikely to contribute to fitness. Because our populations are asexual it is possible for deleterious  
213 mutations to fix by hitchhiking on the background of strong driver mutations. We were surprised,  
214 however, to find that deleterious mutations are nearly absent from our dataset. Only three of 116  
215 mutations were identified as deleterious, two of which are weak mutations (-0.6%) that could represent  
216 false positives. The lack of deleterious mutations is in contrast to an earlier study that found frequent  
217 hitchhiking of deleterious mutations in similar populations (12). We note that our use of 95% confidence  
218 intervals rather than standard error is more conservative, and may account for most of the discrepancy.  
219 We do have compelling evidence for at least one significant deleterious mutation: a stop codon  
220 readthrough of *gcn2* in population BYS1D08-1000. This mutation converts a stop codon (TAG) into a  
221 lysine residue (AAG), resulting in a predicted 26-amino acid extension of the C-terminus of the protein.  
222 The background-averaged fitness effect of this mutation is -2.26%, which is likely an underestimate of  
223 its cost since the mutant *gcn2* allele was not detected in generation 90 of the bulk-segregant fitness  
224 assay (out of the 107 reads that mapped to this position). It is unclear from our data if this *gcn2*

225 mutation is deleterious in all backgrounds; therefore, it is possible that this mutation was beneficial on  
226 the background in which it arose.

227 Among the 116 mutations across the eleven evolved lineages in this study, we measured the  
228 fitness effect of multiple evolved alleles of the same gene, such as *ira1*, and multiple mutations in the  
229 same pathway, such as the mating pathway. The fitness effect of alternate adaptive alleles is  
230 remarkably consistent for the *ira1/2*, *gas1*, and *yur1* mutations, all of which differ by less than 1% in  
231 background-averaged fitness effect (Fig. S6). However, there are exceptions. A single neutral *ste12*  
232 mutation (frameshift at position 472) from the RMS1G02-825 lineage falls outside the narrow range  
233 (2.6%-3.3%) of the four adaptive mutations in the mating pathway. Additionally, the neutral *kel1*-Q107K  
234 mutation in the RMS1G02-825 lineage stands in contrast to the net-beneficial *kel1*-P344T mutation  
235 from BYS2E01-745 (Fig. S6), though we have shown that the latter is also neutral except in the  
236 presence of the co-evolved *hsl7* mutation. It is to be expected that established targets of selection  
237 acquire neutral mutations by chance so these evolved alleles may sweep by hitchhiking.

238 While common targets of selection may represent a substantial fraction of realized adaptive  
239 mutations, it is evident in this study and elsewhere (20) that some lineages may owe their success to  
240 the acquisition of a rare beneficial mutation or an assembly of epistatically-interacting mutations. Here,  
241 we identify synergistic epistasis between two mutations, *kel1* and *hsl7*, that arose within a single cohort.  
242 This observation leads to further questions surrounding the prevalence of such epistatic interactions,  
243 the context in which these interactions arise, and the extent to which epistasis, in addition to neutral  
244 genetic hitchhiking, gives rise to cohort dynamics.

245 In addition to quantifying fitness effects and identifying epistatic interactions, our data reveal a more  
246 detailed view of the dynamics of adaptation. Mapping driver and hitchhiker mutations to each of our 31  
247 mutational cohorts shows that most cohorts consist of 1-2 driver mutations and 0-8 hitchhiker  
248 mutations. However, we also observe that approximately one-third of cohorts lack driver mutations.  
249 Visual inspection of the adaptive dynamics of these “driverless cohorts” reveals that their sweeps are  
250 intimately tied to other concurrent sweeps within the lineage. Driverless cohorts owe their success to  
251 hitchhiking, but unlike mutations within a cohort that arise synchronously from low frequency, mutations  
252 within driverless cohorts occur on the background of the preceding cohort during its climb. The ultimate  
253 fate of a driverless cohort is first tied to the sweep of the preceding cohort. If the preceding cohort is  
254 eventually driven to extinction by a competing lineage, the driverless cohort follows it into extinction  
255 (Fig. S4). Alternatively, if the preceding cohort fixes, the driverless cohort will remain at an intermediate  
256 frequency and its fate will be decided by the next selective sweep. We observe five instances of  
257 driverless cohorts fixing (Fig. S4). In each case this is a two-step process where a driverless cohort is



258 first “pulled” by the preceding cohort and then “pushed” by the ensuing cohort. These “pull-push  
259 dynamics” explain a significant proportion, if not all, of the observed driverless cohorts. Based on our  
260 findings, we contend that driverless cohorts are prevalent across our evolution experiment and that the  
261 presence of a cohort does not necessitate the presence of a beneficial mutation. Our data therefore  
262 support three distinct modes of genetic hitchhiking: hitchhikers within a cohort, cohorts of hitchhikers  
263 “pulled” by a preceding driver mutation, and cohorts of hitchhikers “pushed” by a subsequent driver  
264 mutation.

265 Quantification of fitness can be complicated by intricate dependencies and interactions that have  
266 been shown to arise in laboratory evolution experiments due to nutrient crossfeeding or spatial  
267 structuring (28-32). One of our evolved clones (BYS1E03-745) exhibits negative frequency-dependent  
268 fitness when competed against an ancestral reference: the clone plateaued at frequency of 0.75  
269 regardless of starting proportion (Table S1 and Fig. S8). We suspect the *erg11* mutation present in this  
270 clone since defects in ergosterol biosynthesis (or drug-based inhibition of the pathway) have been  
271 shown to result in negative frequency-dependence in yeast (28). We observed abnormal well-  
272 morphology *erg11*-containing cultures, with cells dispersed to the edges in addition to the bottom of the  
273 well, mirroring the morphology of the “adherent” A-type cells described previously. (28). Because our  
274 bulk-segregant pool contained random combinations of evolved mutations, both with and without the  
275 *erg11* mutation, we could successfully quantify the fitness effect of all mutations in this clone with the  
276 exception of the *erg11* allele itself, which maintained a frequency between 0.8 and 0.9 for the entirety of  
277 the 90-generation bulk-segregant fitness assay. Curiously, the *erg11* allele did not impart frequency  
278 dependence according to the adaptive dynamics of the evolving BYS1E03 population (10). Instead,  
279 *erg11* rapidly fixed in the evolution experiment. The *erg11* mutation may have been “pushed” through  
280 the population by the preceding *ira1*-driven cohort. In other words, the *ira1* allele may have masked the  
281 frequency dependence of the *erg11* allele.

282 This study demonstrates the power of experimental evolution to identify epistatic interactions. Much  
283 of our understanding of epistasis in budding yeast comes from the systematic analysis of double  
284 mutants (33). While informative for constructing large-scale genetic interaction networks, these studies  
285 have thus far been restricted to gene amplifications and deletions, which fail to capture a significant  
286 portion of the mutational spectrum. The observed *kel1-hs17* interaction has not been identified by these  
287 systematic approaches, and several lines of evidence suggest that the interacting *hs17* allele results  
288 from a rare mutational event. First, no other *hs17* mutations were detected in any of the other 40 time-  
289 course sequenced populations (10) despite the BYS2E01-*hs17* allele conferring a 2.1% advantage on  
290 the ancestral background. This *hs17* allele encodes a truncated version of the protein that lacks only the

291 C-terminal domain, which is phosphorylated by Hsl1p in cell cycle-dependent manner to relocate Hsl7p  
292 from the spindle-pole body to the bud neck (34). Deletion of *HSL7* is deleterious under a wide-range of  
293 conditions, including the rich glucose media used here (35, 36), thus our data suggests that the evolved  
294 *hsl7* allele bestows a novel function or alters an existing function. Extensive characterization of such  
295 rare beneficial mutations requires long-term, high replicate evolution experiments followed by  
296 comprehensive analysis linking genotype to phenotype. Likely due to their large target size, loss-of-  
297 function mutations dominate adaptive evolution experiments, though rare beneficial mutations and  
298 epistatic interactions may provide the raw material for molecular innovation in natural populations.

299

## 300 **METHODS**

### 301 **Strain construction**

302 We generated a *MAT $\alpha$*  strain with the following genotype: *ade2-1, ura3::CAN1, his3-11, leu2-3,112,*  
303 *trp1-1, can1, bar1 $\Delta$ ::ADE2, hml $\alpha$  $\Delta$ ::LEU2*. This strain (yGIL737) differs from the ancestral strains from  
304 the original evolution experiment (8) in that it is *MAT $\alpha$* , it has a loss-of-function mutation in *can1*, it has  
305 a wild-type copy of *CAN1* at the *ura3* locus, and it does not contain the NatMX cassette linked to *GPA1*,  
306 which is variable between the two ancestral strains (DBY15095/yGIL429 and DBY15092/yGIL432).

307 We crossed our *MAT $\alpha$*  evolved clones to yGIL737 and, when necessary, we complemented sterile  
308 mutations in the evolved clones using the appropriate plasmid from the MoBY ORF plasmid collection  
309 (37). The resulting diploid strains were converted to *MAT $\alpha/a$*  by a 3-hour galactose induction using a  
310 plasmid harboring a Gal-driven HO. *MAT $\alpha/a$*  convertants were identified by the formation of shmoos  
311 following a 3h.  $\alpha$ F-induction and were verified by PCR of the mating-type locus using primers *MAT $\alpha$ -*  
312 *internal-F* (5' GCACGGAATA TGGGACTACT TCG 3'), *MAT $\alpha$ -internal-F* (5' ACTCCACTTC  
313 AAGTAAGAGT TTG 3'), and *MAT-external-R* (5' AGTCACATC AAGATCGTTT ATGG 3'). Each  
314 *MAT $\alpha/a$*  diploid was transformed with a *URA3* plasmid, pGIL071, which harbors the *MAT $\alpha$ 2* locus  
315 needed for sporulation. Note that using the full-length *MAT $\alpha$*  locus produces rare *MAT $\alpha$*  spores,  
316 presumably due to gene conversion at the *MAT* locus. Therefore the *MAT $\alpha$ 1* gene, which is required for  
317 *MAT $\alpha$*  mating but not sporulation, was not included on our plasmid.

### 318 **Generating segregant pools and isolating individual segregants**

319 To create segregant pools, a single colony of each *MAT $\alpha/a$*  diploid containing the plasmid-borne  
320 *MAT $\alpha$ 2* was inoculated into 10 ml of YPD and grown overnight to saturation. Overnight cultures were  
321 spun down and resuspended in 20 mL SPO++ (1.5% Potassium Acetate, 0.25% Yeast Extract, 0.25%  
322 Dextrose, supplemented with 1 x CSM-arg, Sunrise Science) and sporulated on a room temperature

323 roller drum for 3-4 days until ~75% sporulation efficiency was reached. Sporulated cultures were spun  
324 down and resuspended in 200  $\mu$ l H<sub>2</sub>O for a final volume of 500-750  $\mu$ l. Ascus walls were digested by  
325 adding 5  $\mu$ l of zymolyase (150 mg/mL) and incubating for 1h at 30°C. Next, 50  $\mu$ l glass beads and 50  $\mu$ l  
326 10% Triton were added and the asci were disrupted by vortexing for 2 min. This was followed by an  
327 additional 40 min at 30°C and again by vortexing for 2 min. Disrupted asci were brought up to 5 ml with  
328 H<sub>2</sub>O and were sonicated using a microtip sonicator for four seconds at full power. A liquid hold was  
329 performed by inoculating 500  $\mu$ l of the spore prep into 5 mL YPD and incubating for 24 hours at 30°C.

330 Segregants were isolated by plating onto 150 mm petri dishes containing solid BSP medium (CSM-  
331 arg, Yeast Nitrogen Base, 2% Dextrose, with 1 g/L 5FOA, 60 mg/L canavanine, and 100 mg/L  
332 ClonNat). To make segregant pools two milliliters of the spore preparation were plated onto duplicate  
333 BSP plates (resulting in ~10<sup>5</sup> segregants across the two plates). After ~3 days of growth at 30°C, 5 ml  
334 of YPD was added to the first plate and a sterile glass spreader was used to remove yeast from the  
335 surface of the agar plate. This volume (2-3 ml) was transferred to the second plate and the process was  
336 repeated and the liquid was removed with a pipette and transferred to a 5 ml tube. To remove residual  
337 yeast cells, a second plate wash was done with 2.5 ml. This liquid from the second was (1-2 ml) was  
338 added to the same the same 5 ml tube for a total of ~4 ml. To isolate individual segregants, one milliliter  
339 of the spore prep was plated onto duplicate BSP plates. After ~3 days of growth at 30°C 192 colonies  
340 were picked and inoculated into 130  $\mu$ l YPD in two 96-well plates. Plates were grown for 1-2 days at  
341 30°C. All segregants were phenotyped for growth on 5FOA and canavanine, for the absence of growth  
342 on SC-ura, and for mating type. Segregants were stored at -80°C in 15% glycerol.

### 343 **Bulk-segregant fitness assay**

344 For each bulk-segregant pool, seven replicate population were set up in a single 96-well plate. The  
345 bulk-segregant pools were propagated for 100 generations in conditions identical to the original  
346 evolution experiment (8). In brief, 132  $\mu$ l cultures were maintained in YPD at 30°C and transferred to  
347 fresh media every 24 hours at a dilution of 1:1024 – identical conditions under which the populations  
348 were maintained during the evolution experiment. The assay lasted for a total of 10 days,  
349 corresponding to 100 generations. Following each transfer, the remaining culture was pooled across  
350 replicates (for a total volume of ~900  $\mu$ L). Cells were pelleted and washed with 1 mL of H<sub>2</sub>O. Frozen  
351 pellets were stored at -20°C. Genomic DNA was prepared from the frozen pellets using a modified  
352 glass bead lysis method (10).

### 353 **Library preparation and whole-genome sequencing**

354 Multiplexed genomic DNA libraries were prepared using a modified variation of the Nextera protocol  
355 (Baym *et al.* 2015) with the following modification: to conserve Nextera Index primers, the Index PCR  
356 was performed for 8 cycles, followed by amplification of the libraries via a 5-cycle Reconditioning PCR  
357 using Primers P1 (5' AATGATACGG CGACCACCGA 3') and P2 (5' CAAGCAGAAG ACGGCATACGA  
358 3'). Libraries were quantified by NanoDrop and pooled at equal concentration. The multiplexed pool  
359 was excised from an agarose gel (QIAquick® Gel Extraction Kit, Qiagen) to size-select for fragments  
360 between 400-700 bp, and the collected fragments were analyzed by BioAnalyzer on High-Sensitivity  
361 DNA Chip (BioAnalyzer 2100, Agilent). The samples were run on an Illumina HiSeq 2500 sequencer  
362 with 250 bp single-end reads by the Sequencing Core Facility within the Lewis-Sigler Institute for  
363 Integrative Genomics at Princeton University.

### 364 **Sequencing data analysis**

365 Raw sequencing data was split by index using a dual-index barcode splitter (`barcode_splitter.py`)  
366 from L. Parsons (Princeton University). To analyze the bulk-segregant fitness data we first generated a  
367 list of potential mutations that segregated within the bulk-segregant pool using our previous whole-  
368 genome, whole-population time-course sequencing data (10) to identify mutations within the evolved  
369 population at the time of clone isolation (Table S2). For each mutation, we generated four search terms  
370 corresponding to the ancestral and evolved alleles in both forward and reverse orientation. Each search  
371 term consisted of the known mutation as well as ten nucleotides immediately upstream and  
372 downstream of the mutation. In most instances, the search term was specific to a single locus within the  
373 genome. If the initial search term lacked absolute specificity in the reference genome, the search term  
374 was extended by five to ten nucleotides in each direction. For each FASTQ file, the number of reads  
375 containing the search terms was recorded, providing output in the form of ancestral and evolved allele  
376 counts at each locus. The search criteria required read accuracy/precision since only reads that  
377 possessed no mismatches/errors within the immediate vicinity of the mutation were counted. For  
378 comparison, the sequence reads were mapped to a corrected W303 genome (10) using BWA version  
379 0.7.12 (38) with default parameters except “Disallow an indel within INT bp towards the ends” set to 0  
380 and “Gap open penalty” set to 5. Mutations were then called using FreeBayes version v0.9.21-24-  
381 g840b412 (39) with default parameters for pooled samples. The two approaches were often  
382 harmonious (Fig. S9), and any significant discrepancies resulted due to FreeBayes calling the same  
383 mutation under several different call variants, particularly at the end of reads, an issue stemming from  
384 forced alignment to a designated ancestral reference.

385 The fitness effect of each mutation was calculated as the linear regression of the log ratio of allele  
386 frequency ( $N_{\text{evolved}}/N_{\text{ancestral}}$ ) over the 100-generation experiment. Data corresponding to any time point

387 at which either the ancestral or evolved allele was undetected were removed from analysis. Standard  
388 error of the regression and 95% confidence intervals were determined using MATLAB™  
389 (MathWorks®). Mutations were classified as neutral if the confidence interval encompassed zero.  
390 Mutations with confidence intervals entirely above or below zero were characterized as beneficial or  
391 deleterious, respectively.

### 392 **Identification of additional allelic variants segregating in bulk-segregant pools**

393 The genotyped bulk-segregant individuals from the BYS2E01 cross were analyzed for the presence  
394 of previously unrecognized genetic variants – alleles that differed between the *MATa* ancestor and  
395 *MATα* parent, or low frequency mutations present in the evolved clone. VCF files corresponding to all  
396 BYS2E01 individuals were merged and scanned for calls that fit the following criteria: 1) the mutation is  
397 only called in a fraction of individuals, 2) the mutation, when called, is present near 100% in most  
398 clones, 3) the mutation is called at less than 100% in the individuals acknowledged to be non-clonal  
399 mixed samples, and 4) the mutation does not have another allelic variant. A total of seven genetic  
400 variants were identified and the sequencing data from the bulk-segregant pool fitness assay were  
401 analyzed using the search term approach (Fig. S10).

### 402 ***IQG1* allele replacement**

403 All individuals in the BYS2E01-745 segregant pool contain the evolved allele of *iqg1*. To determine  
404 if the *iqg1* allele imparts a fitness effect, we replaced the *iqg1* mutation with the wild-type *IQG1* allele in  
405 82 of the 192 segregant individuals. We synthesized a gBlock® gene fragment (Integrated DNA  
406 Technologies) containing a guide RNA (AGAAAATATTATGAAGTTTT) targeting the *iqg1* mutation and  
407 the adjacent PAM sequence (CGG). The fragment was first cloned into the *URA3*-marked plasmid  
408 p426-SNR52p-gRNA-SUP4t (Addgene #43803). We then inserted the *HIS3* gene into the *BtgI*  
409 restriction sites in *URA3* creating a partial *ura3* deletion. The resulting *HIS3* gRNA plasmid, a *TRP1*-  
410 marked plasmid containing a constitutively expressed Cas9 (Addgene #43802), and a *URA3*-marked  
411 plasmid containing the wild-type *IQG1* allele (37) were co-transformed into 82 *iqg1* segregants.  
412 Transformants were selected on media utilizing the aforementioned auxotrophic markers, and all  
413 plasmids were eventually cured following verification of the allele swap. To screen for convertants, we  
414 utilized the *SspI* restriction site that was introduced through the evolved *iqg1* mutation (AGTATT →  
415 AATATT). Following PCR and *SspI* digest, transformants that yielded an intact 505 bp PCR product  
416 were presumed to be *IQG1* convertants. We performed Sanger sequencing on a subset of converted  
417 clones (6) to confirm the allele swap and ensure that no other mutations were introduced into the 505  
418 bp region around the Cas9 cut site. Overall efficiency of allele replacement ranged from 15% to 85%.

## 419 **Flow cytometry-based competitive fitness assays**

420 Fitness of the evolved clones and bulk-segregant individuals was determined in flow cytometry-  
421 based competitions as described previously (8, 10). Briefly, segregants were mixed 1:1 with a  
422 ymCitrine-labelled ancestral reference and passaged in YPD broth at 1:1024 dilution on a 24-hr cycle to  
423 mimic evolution conditions. Every 10 generations, an aliquot was transferred to PBS and assayed by  
424 flow cytometry (BD FACSCanto™ II). Cytometric data was analyzed by FlowJo™. Fitness was  
425 calculated as the linear regression of the log ratio of experimental-to-reference frequencies over the 40-  
426 generation assay. Extremely low-fitness clones, which presumably acquired a deleterious mutation,  
427 were not included in downstream analysis of the aggregate data. The clone isolated from population  
428 BYS1E03 displays negative frequency dependent selection due to a mutation in *erg11*. Mutations in the  
429 ergosterol pathway or chemical inhibition of Erg11p, give rise to spatial heterogeneity and negative  
430 frequency-dependent selection in our system (28).

431

## 432 **ACKNOWLEDGEMENTS**

433 We thank Michael Desai, Andrew Murray, and members of the Lang Lab for their comments on the  
434 manuscript. This work was supported by The Charles E. Kaufman Foundation of The Pittsburgh  
435 Foundation.

436

## 437 **FIGURE LEGENDS**

438 **Fig. 1: Genetic dissection of mutations from two evolved lineages.** (A) Genome evolution of each  
439 population was previously tracked through time-course, whole-genome sequencing (10). An evolved  
440 clone was isolated from each population at defined time points. Each trajectory represents a unique  
441 mutation within the isolated clone (colored by chromosome), whereas gray trajectories indicate  
442 mutations detected in competing lineages within a population. (B) The background-averaged fitness  
443 effect of each evolved mutation is measured through a bulk-segregant fitness assay where a segregant  
444 pool is propagated in the selective environment and allele frequencies are tracked by whole-genome  
445 time-course sequencing. Fitness is calculated as the linear regression of the natural log ratio of evolved  
446 to ancestral allele frequency over time. The color scheme remains consistent between the evolutionary  
447 trajectories and bulk-segregant fitness assay. The dynamics of each mutation during the evolution  
448 experiment and the bulk-segregant fitness assay are in Fig. S1 and Table S2. (C) Individual clones  
449 isolated from a bulk-segregant pool are assayed for fitness against an ancestral reference in a flow  
450 cytometry-based competition to determine how fitness segregates in the cross (yellow). The fitness



451 distribution of segregants derived from an ancestral cross (top panel: BY, bottom panel: RM) provides a  
452 baseline fitness in the absence of beneficial alleles. The fitness distribution of the individual segregants  
453 is compared to the fitness of the evolved clone from which they arose (green). The fitness for all 192  
454 segregants from each of the eleven lineages is available in Table S3.

455 **Fig. 2: Comprehensive quantification of fitness effects.** Background-averaged fitness effects for  
456 116 evolved mutations are quantified by bulk-segregant fitness assay. Mutations are separated by  
457 lineage of origin and ordered by mean fitness effect. Fitness effect is represented as the standard error  
458 of regression (thick lines) and 95% confidence interval (thin lines). Mutations are considered neutral  
459 (blue) if the confidence interval encompasses zero. Beneficial (orange) and deleterious (purple)  
460 mutations possess confidence intervals that fall completely above or below zero, respectively.

461 **Fig. 3: Mutational signatures, cohort composition, and additivity of fitness effects.** A) Mutations  
462 were divided into categories based upon their protein coding effect. The mutational signature of driver  
463 mutations is distinct from that of hitchhiker mutations ( $p < 10^{-9}$ ; Chi-square). B) Hierarchical clustering  
464 identified 31 cohorts amongst the eleven evolved lineages. Cohorts vary in size from one to ten  
465 mutations and contain between zero and two drivers (Fig. S3). We observe a positive relationship  
466 between the number of drivers within a cohort correlates with cohort size ( $\rho = 0.70$ ; Pearson correlation).  
467 C) Fitness of all eleven evolved clones correlates with the sum of the fitness effects of their underlying  
468 evolved mutations, as quantified through bulk-segregant fitness assay. Vertical error bars reflect the  
469 standard error between replicate competitions of a common clone, and horizontal error bars reflect the  
470 propagation of error corresponding to the summation of individual background-averaged fitness effects.  
471 Deviation from the dashed line indicates non-additive genetic interactions. The BYS2E01-745 clone  
472 (green) deviates furthest from the expectation.

473 **Fig. 4: Adaptation mechanisms include rare mutations and epistatic interactions.** A) Evolutionary  
474 dynamics of population BYS2E01 as tracked through whole-genome time-course sequencing (10). A  
475 beneficial *ste12* mutation (gray) outcompeted an 11-member cohort (colored) that is enriched for  
476 mutations in genes whose protein products localize to the cellular bud and site of polarized growth. B)  
477 Bulk-segregant individuals from the BYS2E01-745 cross were genotyped and assayed for fitness,  
478 producing a genotype-to-phenotype map. Two evolved alleles, *hsl7* and *kel1*, are associated with  
479 fitness gain. Shown are the average fitness and standard deviation of segregants when parsed by  
480 *HSL7* and *KEL1* alleles. The *kel1* mutation only confers a benefit in the *hsl7* background, and the  
481 fitness advantage of the *hsl7 kel1* double mutant is greater than the sum of the *hsl7* and *kel1* single  
482 mutants.

483 **Fig. S1 Genetic dissection of mutations from all eleven evolved clones.** (A) Genome evolution of  
484 each population was previously tracked through time-course, whole-genome sequencing (10). An  
485 evolved clone was isolated from each population at defined time points. Each trajectory represents a  
486 unique mutation, colored by chromosome, within the isolated clone, whereas gray trajectories indicate  
487 mutations detected in competing lineages within a population. (B) The background-averaged fitness  
488 effect of each evolved mutation is measured through a bulk-segregant fitness assay where a segregant  
489 pool is propagated in the selective environment and allele frequencies are tracked by whole-genome,  
490 time-course sequencing. Fitness is calculated as the linear regression of the natural log ratio of evolved  
491 to ancestral allele frequency over time. The color scheme remains consistent between the evolutionary  
492 trajectories and bulk-segregant fitness assay. (C) Individual clones isolated from a bulk-segregant pool  
493 are assayed for fitness against an ancestral reference in a flow cytometry-based competition to  
494 determine how fitness segregates in the cross (yellow). The fitness distribution of the individual  
495 segregants is compared to the fitness of the evolved clone from which they arose (green). Shown is the  
496 additive expectation of all driver mutations (blue) and all detected mutations (red) as measured by bulk-  
497 segregant fitness assay.

498 **Fig. S2: Construction of bulk-segregant pools and implementation of bulk-segregant fitness**  
499 **assays.** A) To uncover how evolved mutations impact fitness, an evolved clone is selected for bulk-  
500 segregant analysis. First, the evolved *MATa* clone is crossed to a *MATa* version of the ancestor  
501 engineered to possess a functional *CAN1* gene and non-functional *can1* gene at unlinked loci. The  
502 resulting *MATa/a* diploid is heterozygous at all loci mutated in the evolved clone. The diploid is then  
503 mate-type switched to *MATa/a* to ensure that only *MATa* progeny are produced. To sporulate, the  
504 diploid is complemented with a plasmid harboring the *MATa2* gene, and isolation on canavanine-  
505 containing media selects against any unsporulated diploids. Each segregant contains a random  
506 combination of evolved mutations, and the pool, collectively, contains all possible genotypes. B) Each  
507 bulk-segregant pool is propagated in the selective environment for 100 generations, during which time  
508 allele frequencies change based on their fitness effect. Samples are collected every 20 generations and  
509 analyzed by whole-genome sequencing. C) The background-averaged fitness effect of a mutation is  
510 measured as the change in the natural log ratio of allele frequency over time. Beneficial mutations (red  
511 and blue) increase in frequency while neutral mutations (orange) remain steady over time.

512 **Fig. S3: Mutational cohort clustering and dynamics of selected evolved lineages.** Mutations were  
513 assigned into cohorts through grouping of evolved alleles based on a hierarchical clustering approach  
514 that leverages the known evolutionary dynamics. The heat maps reflect allele frequencies, and

515 dendrograms display the distance and relationship between mutations. Mutations are colored by cohort  
516 designation, in accordance with both the dendrogram and evolutionary trajectories.

517 **Fig. S4: Driverless cohorts sweep due to their genetic association with neighboring cohorts.**

518 Comprehensive analysis of evolved alleles revealed that a significant number of mutational cohorts do  
519 not possess any driver mutations. These driverless cohorts (red) often contain few mutations and  
520 exhibit evolutionary trajectories that mimic the preceding and ensuing driver-containing cohorts (gray).  
521 A driverless cohort can be pulled up by a previous cohort (e.g. Cohort B pulled by Cohort A in  
522 population RMS1G02, non-sterile clone), pushed up by an ensuing cohort (e.g. Cohort A pushed by  
523 Cohort B in population RMS1D12), or a combination of both (e.g. Cohort B pulled by Cohort A then  
524 pushed by Cohort C in population BYS1D08). The observed evolutionary dynamics of driverless  
525 cohorts can be explained without invoking adaptive mechanisms.

526 **Fig. S5: Sequence confirmation of the gene conversion event at the *IQG1* locus during**  
527 **construction of the BYS2E01 bulk-segregant pool.** To generate the BYS2E01 bulk-segregant pool,  
528 the evolved clone (*iqg1*) was crossed to the ancestor (*IQG1*). The resulting *MATa/α* diploid exhibited  
529 the expected *IQG1/iqg1* genotype. However, after mate-type switching, the resultant *MATa/a* diploid  
530 possessed only the *iqg1* allele, indicative of a gene conversion event. Following sporulation of the  
531 *MATa/a* diploid, all haploid segregants possessed the *iqg1* allele.

532 **Fig. S6: Evolved mutations in common targets of selection confer similar fitness effects.** We  
533 compared the background-averaged fitness effects of mutations in common targets of selection,  
534 defined as genes or genetic pathways, which were mutated more than once within the eleven evolved  
535 lineages in this study. Thick bars refer to the standard error and thin bars refer to the 95% confidence  
536 interval of the background-averaged fitness effect of each mutation as determined through propagation  
537 and sequencing of each bulk-segregant pool (see Fig. 2). Both driver mutations (orange) and neutral  
538 mutations (blue) are represented. Each mutation is labeled by its predicted coding change (FS:  
539 frameshift, \*: stop codon).

540 **Fig. S7: Fitness effect alone fails to predict the number of occurrences of each driver mutation.**

541 The number of observed mutations in a given gene across forty replicate populations from Ref. (10) is  
542 not correlated to its background-average fitness effect ( $\rho = 0.11$ ,  $p = 0.66$ , Spearman, two-tailed). For  
543 *gas1*, *kre6*, and *yur1*, the fitness values are the average of two measurements from independent  
544 populations. Similarly, for *ira1* the fitness value is the average of four measurements from independent  
545 populations. For *kel1* the fitness value is only from clone BYS2E01-745 and does not include the  
546 neutral *kel1* mutation in population clone RMS1G02-825. Open circles are driver mutations that were  
547 missed by our previous recurrence-based methods that identified putative driver mutations based on

548 three or more observations across forty replicate populations (14). These include modest-effect  
549 mutations (~1-2%) as well as the large-effect interacting mutations, *kel1* and *hsl7*.

550 **Fig. S8: Evolved clone from population BYS1E03 exhibits negative frequency-dependent fitness.**

551 Each evolved clone was assayed for fitness against an ancestral reference strain in seven replicate  
552 competitions. (A) In each competition, the evolved clone was seeded at four different starting  
553 frequencies (purple: 25%, green: 50%, orange: 75%, and red: 90%). The BYS2D06-910 competitions  
554 converge to a common ratio regardless of initial frequency, and subsequently its fitness (i.e. slope)  
555 decreases over time. In contrast, the fitness of BYS1E03-745 and all other evolved clones (Table S1)  
556 remains constant over time. (B) The instantaneous fitness values of clones BYS2D06-910 and  
557 BYS1E03 are calculated as the change in the log ratio of allele frequency across two consecutive  
558 timepoints ( $n$  and  $n+10$ ). The fitness of clone BYS2D06 is inversely proportional to competition  
559 frequency ( $\rho=-0.72$ ; Pearson correlation), whereas all other clones are independent of frequency,  
560 including BYS1E03 shown here ( $\rho=0.10$ , Pearson correlation).

561 **Fig. S9: Comparison of methods for determining allele counts within the bulk-segregant fitness**

562 **assay.** Allele frequencies for the bulk-segregant fitness assay were determined through the analysis of  
563 Illumina sequencing data using two independent approaches: 1) BWA/FreeBayes and 2) our in-lab  
564 'search term' approach. In both cases, the fitness effect of each evolved allele was calculated as the  
565 change in log ratio of allele frequency over time. Data from the first round of sequencing, which  
566 represents ~50% of total read coverage, was used to compare approaches. Manually investigation of  
567 the discrepancies validated the 'search term' approach, which was then used exclusively for all  
568 downstream analysis.

569 **Fig. S10: Newly discovered allelic variants have minimal impact on fitness.** Previously unknown

570 allelic variants were identified through sequencing of the bulk-segregant individuals from the BYS2E01  
571 cross. The frequencies of the seven discovered allelic variants were then tracked over time within all  
572 eleven bulk-segregant pools. Most variants are present in all pools, indicating a disparity between the  
573 *MATa* ancestor and *MAT $\alpha$*  parent. An allele detected in only a fraction of the eleven pools (i.e.  
574 chrII\_625980) was likely present at an intermediate frequency in the *MATa* ancestral or *MAT $\alpha$*  parental  
575 culture. An allele specific to progeny from a single cross (i.e. chrIV\_593410) was likely a low-frequency  
576 evolved mutation. Most genetic variants appear neutral by the bulk-segregant fitness assay. Variants  
577 that depart from neutral are explained via linkage to a known driver mutation.

578

579

580

581 **REFERENCES**

582

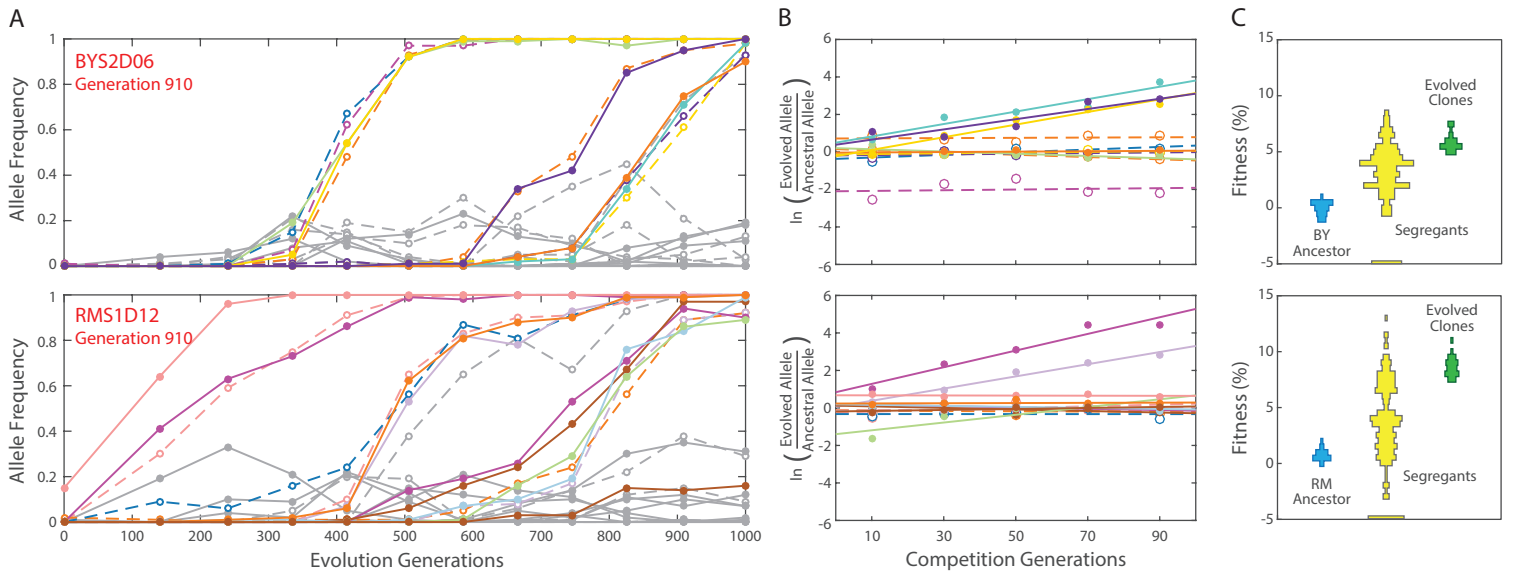
- 583 1. Payen C, *et al.* (2016) High-Throughput Identification of Adaptive Mutations in Experimentally  
584 Evolved Yeast Populations. *PLoS genetics* 12(10):e1006339.
- 585 2. Venkataram S, *et al.* (2016) Development of a Comprehensive Genotype-to-Fitness Map of  
586 Adaptation-Driving Mutations in Yeast. *Cell* 166(6):1585-1596.e1522.
- 587 3. Atwood KC, Schneider LK, & Ryan FJ (1951) Periodic selection in *Escherichia coli*. *Proceedings*  
588 *of the National Academy of Sciences of the United States of America* 37(3):146-155.
- 589 4. Paquin C & Adams J (1983) Frequency of fixation of adaptive mutations is higher in evolving  
590 diploid than haploid yeast populations. *Nature* 302(5908):495-500.
- 591 5. Bryson V & Szybalski W (1952) Microbial Selection. *Science (New York, N.Y.)* 116(3003):45-51.
- 592 6. Gerrish PJ & Lenski RE (1998) The fate of competing beneficial mutations in an asexual  
593 population. *Genetica* 102-103(1-6):127-144.
- 594 7. Kao KC & Sherlock G (2008) Molecular characterization of clonal interference during adaptive  
595 evolution in asexual populations of *Saccharomyces cerevisiae*. *Nature genetics* 40(12):1499-  
596 1504.
- 597 8. Lang GI, Botstein D, & Desai MM (2011) Genetic variation and the fate of beneficial mutations in  
598 asexual populations. *Genetics* 188(3):647-661.
- 599 9. Perfeito L, Fernandes L, Mota C, & Gordo I (2007) Adaptive mutations in bacteria: high rate and  
600 small effects. *Science (New York, N.Y.)* 317(5839):813-815.
- 601 10. Lang GI, *et al.* (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving  
602 yeast populations. *Nature* 500(7464):571-574.
- 603 11. Maddamsetti R, Lenski RE, & Barrick JE (2015) Adaptation, Clonal Interference, and  
604 Frequency-Dependent Interactions in a Long-Term Evolution Experiment with *Escherichia coli*.  
605 *Genetics* 200(2):619-631.
- 606 12. McDonald MJ, Rice DP, & Desai MM (2016) Sex speeds adaptation by altering the dynamics of  
607 molecular evolution. *Nature* 531(7593):233-236.
- 608 13. Moura de Sousa JA, Alpedrinha J, Campos PR, & Gordo I (2016) Competition and fixation of  
609 cohorts of adaptive mutations under Fisher geometrical model. *PeerJ* 4:e2256.
- 610 14. Lang GI, Murray AW, & Botstein D (2009) The cost of gene expression underlies a fitness trade-  
611 off in yeast. *Proceedings of the National Academy of Sciences of the United States of America*  
612 106(14):5755-5760.
- 613 15. Tong AH, *et al.* (2004) Global mapping of the yeast genetic interaction network. *Science (New*  
614 *York, N.Y.)* 303(5659):808-813.
- 615 16. Lang GI & Desai MM (2014) The spectrum of adaptive mutations in experimental evolution.  
616 *Genomics* 104(6 Pt A):412-416.
- 617 17. Boyle EI, *et al.* (2004) GO::TermFinder--open source software for accessing Gene Ontology  
618 information and finding significantly enriched Gene Ontology terms associated with a list of  
619 genes. *Bioinformatics (Oxford, England)* 20(18):3710-3715.
- 620 18. Quandt EM, Deatherage DE, Ellington AD, Georgiou G, & Barrick JE (2014) Recursive  
621 genomewide recombination and sequencing reveals a key refinement step in the evolution of a

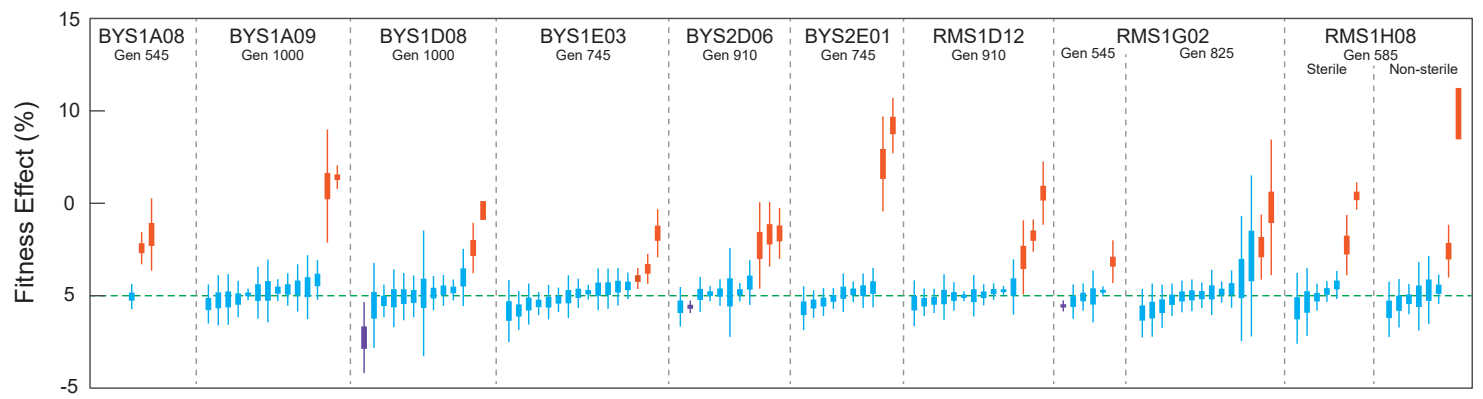


- 622 metabolic innovation in *Escherichia coli*. *Proceedings of the National Academy of Sciences of*  
623 *the United States of America* 111(6):2217-2222.
- 624 19. Quandt EM, *et al.* (2015) Fine-tuning citrate synthase flux potentiates and refines metabolic  
625 innovation in the Lenski evolution experiment. *eLife* 4.
- 626 20. Blount ZD, Borland CZ, & Lenski RE (2008) Historical contingency and the evolution of a key  
627 innovation in an experimental population of *Escherichia coli*. *Proceedings of the National*  
628 *Academy of Sciences of the United States of America* 105(23):7899-7906.
- 629 21. Chou HH, Berthet J, & Marx CJ (2009) Fast growth increases the selective advantage of a  
630 mutation arising recurrently during evolution under metal limitation. *PLoS genetics*  
631 5(9):e1000652.
- 632 22. Hong J & Gresham D (2014) Molecular specificity, convergence and constraint shape adaptive  
633 evolution in nutrient-poor environments. *PLoS genetics* 10(1):e1004041.
- 634 23. Kvittek DJ & Sherlock G (2013) Whole genome, whole population sequencing reveals that loss  
635 of signaling networks is the major adaptive strategy in a constant environment. *PLoS genetics*  
636 9(11):e1003972.
- 637 24. Lieberman TD, *et al.* (2011) Parallel bacterial evolution within multiple patients identifies  
638 candidate pathogenicity genes. *Nature genetics* 43(12):1275-1280.
- 639 25. Tenaillon O, *et al.* (2012) The molecular diversity of adaptive convergence. *Science (New York,*  
640 *N.Y.)* 335(6067):457-461.
- 641 26. Bailey SF & Bataillon T (2016) Can the experimental evolution programme help us elucidate the  
642 genetic basis of adaptation in nature? *Molecular ecology* 25(1):203-218.
- 643 27. Brockhurst MA (2015) Experimental evolution can unravel the complex causes of natural  
644 selection in clinical infections. *Microbiology (Reading, England)* 161(6):1175-1179.
- 645 28. Frenkel EM, *et al.* (2015) Crowded growth leads to the spontaneous evolution of semistable  
646 coexistence in laboratory yeast populations. *Proceedings of the National Academy of Sciences*  
647 *of the United States of America* 112(36):11306-11311.
- 648 29. Friesen ML, Saxer G, Travisano M, & Doebeli M (2004) Experimental evidence for sympatric  
649 ecological diversification due to frequency-dependent competition in *Escherichia coli*. *Evolution;*  
650 *international journal of organic evolution* 58(2):245-260.
- 651 30. Levin BR (1972) Coexistence of two asexual strains on a single resource. *Science (New York,*  
652 *N.Y.)* 175(4027):1272-1274.
- 653 31. Rainey PB & Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature*  
654 394(6688):69-72.
- 655 32. Rosenzweig RF, Sharp RR, Treves DS, & Adams J (1994) Microbial evolution in a simple  
656 unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* 137(4):903-917.
- 657 33. Baryshnikova A, Costanzo M, Myers CL, Andrews B, & Boone C (2013) Genetic interaction  
658 networks: toward an understanding of heritability. *Annual review of genomics and human*  
659 *genetics* 14:111-133.
- 660 34. Cid VJ, Shulewitz MJ, McDonald KL, & Thorner J (2001) Dynamic localization of the Swe1  
661 regulator Hsl7 during the *Saccharomyces cerevisiae* cell cycle. *Molecular biology of the cell*  
662 12(6):1645-1669.
- 663 35. Brown JA, *et al.* (2006) Global analysis of gene function in yeast by quantitative phenotypic  
664 profiling. *Molecular systems biology* 2:2006.0001.

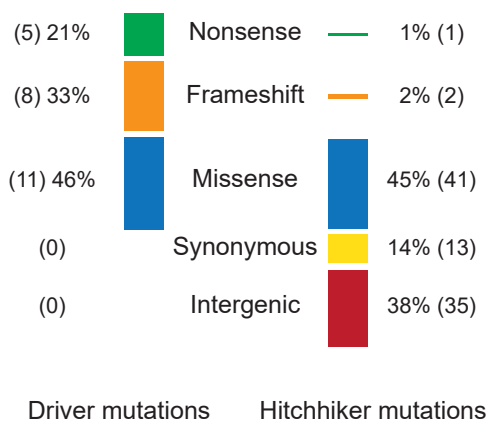


- 665 36. Qian W, Ma D, Xiao C, Wang Z, & Zhang J (2012) The genomic landscape and evolutionary  
666 resolution of antagonistic pleiotropy in yeast. *Cell reports* 2(5):1399-1410.
- 667 37. Ho CH, *et al.* (2009) A molecular barcoded yeast ORF library enables mode-of-action analysis  
668 of bioactive compounds. *Nature biotechnology* 27(4):369-377.
- 669 38. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler  
670 transform. *Bioinformatics (Oxford, England)* 25(14):1754-1760.
- 671 39. Garrison E & Marth G (2012) Haplotype-based variant detection from short-read sequencing.  
672 *arXiv preprint arXiv:1207.3907*.





A



B



C

