Evolutionary dynamics of *de novo* mutations and mutant lineages arising in a simple, constant environment

Margie Kinnersley[1,¶], Katja Schwartz[3,¶], Jacob Boswell[2], Dong-Dong Yang[2], Gavin Sherlock[3,*]
Frank Rosenzweig[1,2,*]

[1] Division of Biological Sciences, The University of Montana, Missoula, MT 59812
[2] School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332
[3] Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305-5120

¶ These authors contributed equally to this work.

[*]Co-corresponding authors: frank.rosenzweig@biology.gatech.edu and gsherloc@stanford.edu

Keywords: *E. coli,* adaptive evolution, chemostat, metagenomics, population sequencing, clone sequencing, clonal interference, parallelism

Running Title: Parallelism and clonal interference in evolving bacterial populations

Abstract

A large, asexual population founded by a single clone evolves into a population teeming with many, whether or not its environment is structured, and whether or not resource levels are constant or fluctuating. The maintenance of genetic complexity in such populations has been attributed to balancing selection, or to either clonal interference or clonal reinforcement, arising from antagonistic or synergistic interactions, respectively. To distinguish among these possibilities, to identify targets of selection and establish when and how often they are hit, as well as to gain insight into how *de novo* mutations interact, we carried out 300-500 generation glucose-limited chemostat experiments founded by an *E. coli* mutator. To discover all *de novo* mutations reaching ≥1% frequency, we performed whole-genome, whole-population sequencing at ~1000X-coverage every 50 generations. To establish linkage relationships among these mutations and depict the dynamics of evolving lineages we sequenced the genomes of 96 clones from each population when allelic diversity was greatest. Operon-specific mutations that enhance glucose uptake arose to high frequency first, followed by global regulatory mutations. Late-arising mutations were related to energy conservation as well as to mitigating pleiotropic effects wrought by earlier regulatory changes. We discovered extensive polymorphism at relatively few loci, with identical mutations arising independently in different lineages, both between and within replicate populations. Out of more than 3,000 SNPs detected in nearly 1,800 genes or intergenic regions, only 17 reached a frequency ≥ 98%, indicating that the evolutionary dynamics of adaptive lineages was dominated by clonal interference. Finally, our data show that even when mutational input is increased by an ancestral defect in DNA repair, the spectrum of beneficial mutations that reach high frequency in a simple, constant resource-limited environment is narrow, resulting in extreme parallelism where many adaptive mutations arise but few ever go to fixation.

Author Summary

Microbial evolution experiments open a window on the tempo and dynamics of evolutionary change in asexual populations. High-throughput sequencing can be used to catalog *de novo* mutations, determine in which lineages they arise, and assess allelic interactions by tracking the fate of those lineages. This *adaptive genetics* approach makes it possible to discover whether clonal interactions are antagonistic or synergistic, and complements genetic screens of induced deleterious/loss-of-function mutants. We carried out glucose-limited chemostat experiments founded by an *E. coli* mutator and performed whole-genome, whole-population sequencing on 300-500 generation evolutions, cataloging 3,346 *de novo* mutations that reached ≥1% frequency. Mutations enhancing glucose uptake rose to high frequency first, followed by global regulatory changes that modulate growth rate and limiting resource assimilation, then by mutations that favor energy conservation or mitigate pleiotropic effects of earlier regulatory changes. We discovered that a few loci were highly polymorphic, with identical mutations arising independently in different lineages, both between and within replicate populations. Thus, when mutational input is increased by an ancestral defect in DNA repair, the spectrum of beneficial mutations that arises under constant resource-limitation is narrow, resulting in extreme parallelism where many adaptive mutations arise but few ever become fixed.

Introduction

Evolution experiments using microbes have enlarged our understanding of the tempo and dynamics of evolutionary change, as well as how selection, drift and historical contingency influence evolutionary trajectories. Combined with high throughput sequencing, experimental microbial evolution (EME) can now be used to identify substantial numbers of *de novo* beneficial mutations in laboratory populations, to determine in which lineages they arise and the fate of those lineages, and to evaluate the sign and strength of possible epistatic interactions [1-3]. This approach, adaptive genetics, based on analyzing cohorts of spontaneous beneficial mutations to determine how their frequencies fluctuate over time, constitutes a mode of inquiry that complements traditional genetic screening of induced deleterious/loss-of-function mutants (e.g., [4] and [5] among others). Adaptive genetics also expands the possibilities for discovering constraints on protein structure and function and for discerning the architecture and malleability of networks that regulate nutrient-sensing and cell division.

Microbial populations were once thought to evolve by periodic selection as a succession of adaptive clones, each fitter than its antecedent, replacing one another over time [6-9]. This model was consistent with Muller and Haldane's view of how beneficial mutations spread in large asexual populations [10-12] under conditions governed by competitive exclusion [13]. Today we know that large, initially clonal populations rapidly accumulate and retain genetic variation, much of which is beneficial [14-18]. In fact, the amount of adaptive genetic variation observed in EME populations can be enormous, owing large population sizes with a continuous input of beneficial mutations and the subsequent competition among new adaptive lineages, which gives rise to clonal interference [14,16,19,20].

Clonal interference can occur within a larger framework of stable subpopulation structure [21] when microbial lineages come under balancing selection [22-25] or specialize to exploit niches created either by the culture conditions [23,26,27], or by the organisms themselves [28-30]. In a simple constant environment like a chemostat the persistence of subpopulations likely depends on founder genotype, the emergence of specific key mutations, and availability of the limiting nutrient [31]. Ferenci and colleagues never observed stable subpopulation structure in glucose-limited evolutions originating from *E. coli* K12 strain BW2952 [32], whereas Adams and colleagues, using a different strain, often did [30,33]. Unlike BW2952, the K12-derived ancestor used by Adams, JA122 [30] harbors a supE44 *glnX* tRNA nonsense suppressor as well as nonsense mutations in housekeeping and stationary-phase transcription factors, RpoD and RpoS respectively, and mismatch repair enzyme MutY. The JA122 ancestor's defect in DNA repair increases mutational load on its descendants [28,34], while the nonsense suppressor mitigates the effect of mutations that create premature stop codons. Such a suppressor would likely make the blunt instrument of *de novo* nonsense mutations a less effective agent of adaptive change, possibly resulting in a more nuanced spectrum of beneficial mutations than would otherwise occur among mutators.

To understand the impact that a mutator/suppressor founder has on the spectrum and fate of new beneficial mutations, and on the dynamics of population structure, we repeated Adams *et al*. classic evolution experiments using the same ancestral strain and culture conditions [30]. We monitored, at 50-generation intervals, the incidence of mutations that reached at least 1% frequency over the course of 300-500 generations, identifying mutations that were either transiently beneficial or hitch-hiking with mutations that were. To determine which mutations co-occurred within a given lineage we sequenced 96 clones from each population at the time-point

where we observed greatest allelic diversity. We uncovered no evidence for stable sub-population structure, but instead saw pervasive clonal interference, with only 17 out of 3,346 mutations going to near fixation across replicate experiments. The temporal order in which certain mutations rose to high frequency was predictable, reflecting a high degree of parallelism both within and between replicates. In general, mutations that enhanced glucose assimilation arose early, followed by mutations in global regulators and mutations that either increase efficiency of limiting resource utilization or mitigate the deleterious effects of certain earlier mutations. Altogether, our results show that even in bacterial populations founded by an ancestor having a high mutation rate and the capacity to tolerate many *de novo* mutations, the spectrum of genomic changes that rise to appreciable frequency and the adaptive outcome of replicate evolutions are limited when those populations evolve in a simple constant environment.

Results

Experimental design. Evolution experiments were carried out in triplicate under continuous nutrient limitation using Davis Minimal Medium [30], with glucose (0.0125% w/v) as the sole source of carbon for energy and growth. Chemostats (300 mL working volume) were run under aerobic conditions for 300-500 generations at constant temperature (30°C) and at constant dilution rate (D=0.2 hr^{-1}). Under these conditions, population density reaches ~10^8 cells mL^{-1} at steady state. The *E. coli* strain used to initiate these experiments, JA122, is distinguished from *E. coli* K12 by alleles likely to influence the spectrum of mutations arising during adaptive evolution (**Table S1**; [28]). Among these is a nonsense mutation in MutY (Leu299*) that results in a 10-fold greater mutation rate and GC→TA transversion bias [28], nonsense mutations in the genes that encode stationary phase sigma factor RpoS (Gln33*) [35] and 'housekeeping' sigma factor RpoD

131 (Glu26*), as well as a suppressor mutation in the *glnX* tRNA known to suppress amber, ochre and
132 opal mutations (**Table S1)** [36].

133 To identify the mutations that arose during the evolutions, we performed whole genome, whole
134 population sequencing every 50 generations on each of the three chemostat populations. We
135 generated approximately 50 million 2x100bp paired end reads per sample, yielding coverage of up
136 to ~1000x for each time point (inserts were selected to be short enough such that forward and
137 reverse reads overlapped, which while reducing coverage, increases quality; see Methods). Based
138 on this level of coverage, we were able to identify mutations that rose to an allele frequency of
139 ~1% of greater. Given an effective population size of $>10^{10}$ and 300-500 generations of selection
140 it is highly improbable that any allele could reach such a frequency by drift alone [23]. We can
141 therefore assume that every mutation recovered was either under positive selection or hitch-hiking
142 along with one that was.

143 ***Population sequencing shows general patterns of mutation that are consistent across***
144 ***independent evolutions***. Across all samples, 3,326 SNPs were detected in 2,083 unique genes or
145 intergenic regions (**File S1)**. The overwhelming majority (97.5%) of these SNPs were GC→TA
146 transversions, as expected given the ancestral strain's defect in the mismatch repair protein MutY,
147 which encodes adenine glycosylase [37]. Consistent with the protein coding density of *E. coli*
148 (87.8%) [38], 85% (2,854) of SNPs occurred in coding regions. On average, 69.2% of these created
149 a missense mutation, 23.4% resulted in a synonymous mutation and 7.4% caused a nonsense
150 mutation (**Fig. 1**). Relative to proportions observed in mutation accumulation experiments carried
151 out using wild-type *E. coli* [39], we observed more nonsynonymous and nonsense mutations.
152 Small deletions were rarely detected (one single-nucleotide deletion in each of vessel 1 and vessel
153 2, and none detected in vessel 3), but we observed a single large ~150kb duplication in vessel 2.

154 The overall number of mutations in each population increased linearly over time and at
155 approximately the same rate across replicates (**Fig. 1**), as would be expected with a mutator
156 phenotype.

157 ***Comparison of population level mutations reveals clonal interference and widespread***
158 ***parallelism.*** Despite the large number of SNPs detected, only 17 alleles arose above a frequency
159 of 98% across replicate evolutions ranging from 300-500 generations. Moreover, the maximum
160 frequency of most alleles never exceeded 10% (**Fig. 2A**), and the vast majority of alleles were
161 present at lower frequency in the final time-point than they were at their maximum (**Fig. 2B**).
162 Together, the foregoing observations suggest that in each evolution experiment population
163 dynamics was largely driven by clonal interference [40]. A small number of loci were recurrently
164 mutated above what would be expected by chance, indicating that variants at these loci were likely
165 beneficial (**Table 1**, **Table S2**). For example, a total of 212 mutations arose in the 10 most
166 significantly mutated genes in the population sequencing data, with each gene receiving at least
167 five mutations (**Table 1**). Moreover, 30 and 14 distinct allelic variants were discovered in just two:
168 the genes encoding the DNA binding repressor GalS and the RNA-binding protein Hfq,
169 respectively (**Table S3**). High-resolution population sequencing also revealed that 13 SNPs not
170 present at the start of the experiment reached at least 1% frequency in all three vessels at various
171 time-points, while 52 SNPs recurred in two out of three chemostats (**Table S4**). Thus, our data
172 also provide compelling evidence for substantial parallel evolution at the genic level.

173 ***Clonal sequencing further clarifies lineage relationships and parallelism*** To determine
174 linkage relationships between the novel alleles, we sequenced 96 individual clones from each
175 vessel. In each case, the 96 clones were isolated at random from the time-point at which we
176 detected the greatest number of mutant alleles at ≥ 5% frequency. To assess whether the frequency

177 estimates from population sequencing were reasonable, and whether the isolated clones constituted
178 a reasonable subsample, we compared frequencies of mutations identified in both datasets at the
179 corresponding time-point and found that they correlate well (**Fig. 3**).

180 For each set of 96 clones, we constructed a phylogeny to represent their putative
181 evolutionary relationships (**Fig. 4**). Inspection of the mutations and trees from each vessel (i.e.
182 each independent evolution) revealed several instances in which exactly the same mutation arose
183 not only in different vessels, but often more than once in the same vessel on distinct branches of a
184 given tree. In the most extreme case, 6 of the 11 *hfq* alleles detected via clone sequencing were
185 identified in clones from different vessels, indicating independent parallel origins (**Fig. 4, File S2**).
186 Furthermore, 7 of the 11 appear to have arisen more than once within the same vessel.

187 ***Clonal dynamics are shaped by relationships among de novo alleles, hard and soft selective***
188 ***sweeps, and absence of periodic selection*** Combining population allele frequency data with
189 linkage information derived from clonal sequencing makes it possible to depict lineage dynamics
190 using Muller diagrams (**Fig. 5, Files S3-S5**). In general, we observe early, hard sweeps of highly
191 beneficial mutations related to limiting nutrient influx, followed by soft sweeps [41-43] and
192 multiple-origin soft sweeps that may fine-tune glucose uptake or utilization later in the experiment
193 when diversity was higher [44-46]. Hard sweeps consistently involved mutations in regulators
194 (*galS* in chemostat 1, transcriptional terminator *rho* in chemostats 1 and 3) or regulatory regions
195 (upstream of *dnaG* in chemostat 1, upstream of *mglB* in chemostats 1, 2 and 3 – See supplementary
196 Files 2, 3 and 4 for detailed dynamics), while soft sweeps were comprised of both regulatory and
197 operon-specific mutations (e.g. *hfq* and *opgH* in chemostats 1, 2 and 3, upstream of *adhE* in
198 chemostat 1, *pgi* in chemostat 3) (**Fig. 5, S1, Files S3-S5)** [42,47]. Here, we note that multiple-
199 origin soft sweeps may be especially prevalent in our experiments due to the ancestral mutator

200 allele at *mutY*, as the likelihood of concurrent identical mutations in the same gene should increase
201 with mutation rate.

202 With regard to periodic selection, rather than favorable alleles arising within a set of lineages
203 that successively replace one another over time, we observe groups or cohorts of mutations co-
204 evolving, with widespread clonal interference among lineages that carry different beneficial
205 mutations [48]. For example, in chemostat 1, a spreading lineage with a cohort of mutations
206 upstream of *mglB*/*lptA*/*opgH* (pink) is checked by the emergence of lineages carrying mutations
207 in *hfq* (green) (**File S3)**. All of these phenomena – hard and soft sweeps, cohorts of mutations that
208 increase or decrease in frequency together, and clonal interference – have been observed in yeast
209 [14,19,49] and *E. coli* [23] populations evolving in the laboratory, as well as in *Pseudomonas*
210 *aeruginosa* evolving in the cystic fibrosis lung [50].

211 ***Early sweeps are related to influx of the limiting nutrient glucose.*** For specific growth rates
212 between $\sim \mu = 0.1$ hr^{-1} and $\mu = 0.9$ hr^{-1}, glucose is most efficiently transported using a combination
213 of the maltoporin LamB and the galactose transporter MglBAC, and glucose limitation frequently
214 selects for mutations that increase expression of these proteins [45,51-59]. As expected, 7 of the
215 top 10 frequently mutated genes/gene regions we observed (*galS*, upstream *mglB*, *malT*, *malK*,
216 *hfq*, *rho* and upstream *dnaG*) play a role in transcriptional regulation of *lamB* or *mglBAC*, either
217 directly or through their interactions with global regulators (**Table 1**, **Fig. 6**).

218 MUTATIONS IN *GALS* AND UPSTREAM OF *MGLB*. Thirty different alleles of *galS* (encoding GalS,
219 a negative regulator of *mglBAC* transcription) were detected over the course of our experiments.
220 These spanned the length of the gene, and the majority caused missense amino acid changes likely
221 to disrupt *mglBAC* transcriptional repression and augment glucose flux across the inner membrane
222 (**Fig. 6**) [60]. Despite the large number of alleles we observed, few persisted beyond generation 50

223 or attained a final frequency greater than 5%, demonstrating high clonal diversity early in the
224 experiment. In chemostats 2 and 3, no clear "winner" *galS* genotype emerged, though in chemostat
225 1, a GalS allele (Arg146Leu) swept to near fixation (89.6% of the population at generation 50).

226 Early-arising *GalS* mutant genotypes were rapidly displaced by clones carrying highly-
227 beneficial mutations in the *mgl* operator sequence upstream of *mglB* (**Fig. 5**). This sequence of
228 events, like the mutations themselves, has been observed elsewhere [45,61]. The most successful
229 mutation upstream of *mglB* (bp 2,238,647 C→A) occurred early in every vessel, and in every case
230 increased in frequency to over 90% of the population (**Table S5**, **File S3-S5**). Notably, this same
231 mutation was observed in chemostat-grown *E. coli* by Notley-McRobb et al. (1999) as well as in
232 the experimental population described by Helling et al. (1987), where it was found to be the only
233 SNP shared by all members of a cross-feeding consortium [34,54].

234 The dynamics of *galS* replacement illustrates the effect that clonal interference can have
235 on the fate of different alleles. In chemostat 1, clones carrying GalS Arg146Leu rapidly dropped
236 in frequency when lineages emerged with a mutation upstream of *mglB* (position 2,238,647), but
237 were not completely displaced until generation 400 and even enjoyed brief periods of expansion.
238 In chemostat 2, clones with the same mutation upstream of *mglB* were present by generation 50,
239 but did not surpass a 90% threshold for another 250 generations due to competition from 22
240 different *galS* lineages and a lineage carrying a different upstream of *mglB* allele (2,238,648 G→T)
241 (**Fig. 5, S1B, File S4**). By contrast, in chemostat 3, a lineage with the upstream *mglB* mutation
242 (2,238,647) experienced little competition and was almost fixed by generation 150 **(Table S5).**

243 Over the remainder of the experiment only three other mutations upstream of *mglB*
244 mutations reached the threshold for detection: two were within 2 base-pairs of the first mutation
245 and did not rise to high frequency, while the third (chemostat 1, 2,238,630 C→A) located in the

246 CRP activator binding site, co-occurred with 2,238,647 C→A and increased to ca. 80% frequency
247 by generation 500 (**Figs. 5, S1A, File S3**). This dynamic suggests additional mutations that affect
248 GalS repressor binding are not of great benefit after the preferred allele has swept the population,
249 whereas mutations that modulate the activity of other regulators (i.e. CRP) can act synergistically.

250 THE DYNAMICS OF LAMB REGULATION. LamB glycoporin overexpression is a hallmark
251 feature of *E. coli* populations adapted to glucose-limited chemostat growth [34,45,53,55,56,62].
252 Previous experiments have shown that under glucose limitation, overexpression of LamB can be
253 the result of any one of the following: constitutive activation of transcriptional regulator MalT,
254 disruption of the MalT inhibitor MalK, mutation of the RNA chaperone Hfq, alteration of sigma
255 factor dynamics (σ^S/ σ^D ratio), or mutation of the *malT* repressor Mlc, (**Fig. 6**) [34,45,53-56,62,63].

256 Across the three replicate evolutions, we observed 19 unique *malT* alleles and 14 unique
257 *malK* alleles (**Figs. 6, S2, Table S3**). Over half of the mutations in *malT* (10 out of 19) are known
258 either to cause MalT to become constitutively active, or to occur in amino acids involved in
259 MalT/MalK interaction [45,64,65]. A single MalK mutation (Ala296Asp) rose to high frequency
260 early (94% by generation 100) in chemostat 1 (**Fig. S2**, **Table S5**). This SNP is in a regulatory
261 domain likely to be the site of MalK/MalT interaction [66]. Alteration of a neighboring residue
262 (Asp297) has been previously shown to allow unregulated transcription of the *mal* operon [66].

263 *malT*/*malK* allele dynamics differ among experimental populations. As mentioned above
264 and shown in **Fig. S2**, MalK Ala296Asp sweeps early in chemostat 1, whereas in chemostat 3,
265 early MalK mutations (blue) are displaced by later mutations in MalT (green). In chemostat 2, the
266 picture is quite different: clones with either *malK* or *malT* mutations co-exist through all 500
267 generations. The reason for this contrast in allele frequency dynamics cannot be attributed to
268 emergence of a single "most fit" allele, as the majority types from chemostats 1 and 3 arose

269 independently in chemostat 2, but did not sweep. Despite the importance of MalT and MalK as
270 high-value targets for selection during adaptation to glucose limitation, other advantageous
271 mutations (upstream *mglB*, *rho* and *hfq*, discussed below) may have ultimately carried "winning"
272 *mal* alleles in chemostats 1 and 3 to higher frequency, purging allelic diversity at this locus.
273 Interestingly, although we observed 30 *malT* and 22 *malK* mutations in the population sequencing
274 data (**Table 1**), in only 5 out of the 288 sequenced clones do mutant alleles of these two genes co-
275 occur, suggesting that there may be no additional advantage or even some disadvantage to having
276 both. In the Helling et al. evolution experiments [30], which were founded by the same ancestor
277 used here, secondary resource specialists share a mutation in MalT, whereas the primary resource
278 (glucose) specialist that feeds those clones carries a mutation in MalK [34].

279 SELECTION OF MUTATIONS IN RNA CHAPERONE HFQ THAT AFFECT TRANSLATION OF LAMB
280 AND STATIONARY PHASE TRANSCRIPTION FACTOR RPOS. Hfq is a global regulatory protein that
281 facilitates translation and/or RNA degradation by mediating ncRNA-mRNA interactions. It
282 participates in a diverse range of cellular processes including nutrient uptake, motility and
283 metabolism [67]. *hfq* mutations identified in other glucose-limited evolution experiments exhibit
284 pleiotropic physiological effects: they appear to increase translation of LamB glycoporin, reduce
285 levels of stationary phase transcription factor RpoS, inhibit cellular aggregation, and enhance
286 glucose transport via PtsG [63,68].

287 *Hfq* is one of the most frequently mutated genes observed in our experiments: 24 *hfq*
288 mutations, resulting in 14 distinct *hfq* alleles, were detected via population sequencing; by the end
289 of our evolutions >50% of each population carried a mutation in *hfq* (**Table S3**). (**Table 1**). Two
290 of these alleles arose independently in all three vessels (same nucleotide position, same SNP), and
291 six additional alleles were observed in two of three vessels (**Table S4**). The frequency of and

292 parallelism exhibited in *hfq* mutations is particularly curious in the context of experiments by
293 Maharjan et al. in which *hfq* mutations arise, but are at low frequency and subject to negative
294 frequency-dependent selection and epistatic interaction with mutations in *rpoS* [32,52,62,63].

295 The dynamics of hfq mutations are variable across evolutions and may depend on which
296 other beneficial alleles are present in the same lineage or in the same population. In chemostats 1
297 and 2, a large number of *hfq* alleles (10 in chemostat 1 and 11 in chemostat 2) appear after
298 generation 250 and are preceded by mutations in *malK* or *malE* and the *opg* operon. The most
299 successful Hfq allele in chemostat 1, Val62Phe, occurred in a sweeping lineage with a secondary
300 mutation upstream of *mglB* (discussed above) and may have been carried along by association. In
301 chemostat 3, a single *hfq* mutation arises early (Ser60Tyr, present by generation 100), sweeps to
302 near fixation alongside MalT Met311Ile and is closely followed by mutations in *opgH* (**Fig. S1,**
303 **File S5**).

304 RECURRENT MUTATIONS OCCUR IN RHO. Early-arising mutations in the *rho* termination factor
305 are a conspicuous feature of chemostats 1 and 3 (**Figs. 5, S1**). Rho is required for transcriptional
306 termination of up to 50% of cellular mRNAs [69,70] and can participate in gene regulation via
307 intragenic terminators [71]. Mutagenesis and ChIP-chip analyses have identified Rho-dependent
308 terminators within multiple genes relevant to glucose limitation, specifically *lamB*, *mglA*, and
309 *mglC* and downstream of *malT* and *mglC* [71,72]. In fact, it has long been known that defective
310 LamB expression in MalT activator mutants can be restored via compensatory mutations in *rho*
311 [73]. In chemostats 1 and 3, Rho mutations fix or nearly fix early and do so in concert with
312 mutations in MalK (chemostat 1 Ala296Asp) and mutations upstream of *mglB* (chemostats 1 and
313 3, bp 2,238,647) (**Figs. 5, S1, Table S3, Table S5, Supplementary Files 2 and 4**). Conversely, in

314 chemostat 2 only three *rho* alleles were detected, none of which rose in frequency to >6% of the
315 population (**Fig. S1, Table S3**).

316 ***Mutations that impact energy conservation, membrane biogenesis and cell adhesion are***
317 ***late arising targets of selection*** PHOSPHOGLUCOSEISOMERASE (PGI) is an abundantly expressed
318 central metabolic enzyme responsible for converting glucose-6-phosphate into fructose-6-
319 phosphate. Knockdown of *pgi* mRNA alleviates catabolite repression [74], favoring increased
320 expression of CRP-regulated genes such as *lamB* and *mglBAC*. Twenty-four unique *pgi* alleles
321 were detected over the course of our three replicate evolutions. However, few rose to appreciable
322 frequency before generation 200, suggesting their benefit may be contingent on the presence of
323 other mutations or some aspect of the chemostat environment that consistently changed after this
324 time point. *Pgi* alleles were least successful in chemostat 1, which was also the only replicate in
325 which a large fraction of clones (79% by generation 500) acquired a second mutation upstream of
326 *mglB*. This observation suggests that *pgi* mutations and mutations in the CRP-binding site of the
327 *mglBAC* promoter may be functionally redundant.

328 MEMBRANE GLYCOSYLTRANSFERASE OPGH is involved in the synthesis of periplasmic glucans,
329 highly branched oligosaccharides made from β-linked glucose monomers. While we do not
330 observe *opgH* mutations earlier than generation 100, they rapidly increase in frequency once they
331 appear, usually either just before or just after *hfq* mutations (**Fig. S1, S4, File S3-S5**). Novel *opgH*
332 alleles, especially the nonsense mutations that we frequently observe, may constrain glucan
333 production and serve as a glucose conservation measure. A "moonlighting" function has also
334 recently been reported for OpgH: the glucosyltransferase interacts with the tubulin-like cell
335 division protein FtsZ to delay cell division when levels of UDP-glucose are low [75]. Thus,
336 mutations in OpgH may augment the rate of cell division, and thereby provide a fitness advantage

337 under slow-growth chemostat conditions. The only *opg* operon mutation identified among strains
338 in previous Adams et al. experiments occurred in *opgG* of the glucose scavenger, CV103 (E487*)
339 [28].

340 MUTATIONS IN RHO-INDEPENDENT TERMINATOR T_1 THAT ALLOW RUN-THROUGH
341 TRANSCRIPTION MAY TIP THE BALANCE BETWEEN COMPETING SIGMA FACTORS. Sigma factor RpoD
342 (σ^{70}) is the predominant sigma factor associated with RNA polymerase during exponential growth.
343 As cells enter stationary phase, transcription of the gene for alternate sigma factor RpoS (σ^S)
344 increases [76]. *rpoS* mutations are often selected for under continuous glucose limitation as they
345 allow continued transcription from promoters negatively regulated by σ^S but required for glucose
346 uptake and metabolism (e.g. [77,78]).

347 In chemostat 1, a mutation in the *rpsU-dnaG-rpoD* macromolecular synthesis operon
348 upstream of *dnaG* (bp 3,209,081 G→T) was present in over 90% of the population by generation
349 50 (**Table S5**). This SNP decreases the stability of the rho-independent terminator T_1 situated
350 between *rpsU* and *dnaG*, and thus may be expected to increase expression of RpoD [79] and as a
351 result operons positively controlled by σ^{70} (e.g. *mglBAC* and *malK-lamB-malM*). A T1 mutation
352 (bp 3,209,075 C→A) was also shared among Helling et al. strains, defining the lineage that gave
353 rise to three of four consortium members [30,34]. In addition, in Chemostat 2, we observed an
354 ~150kb duplication that included *rpoD* and in chemostat 3, eight clones out of 96 carried intragenic
355 suppressor mutations of the ancestral nonsense allele (*26Asp and *26Gln) in RpoD.

356 FIMBRIAL PROTEIN GENES (*FIM*) Genes associated with production/function of type 1 fimbriae,
357 particularly *fimH* (fimbrial adhesion), were an unexpected and frequent target of mutation in all
358 three chemostats (**Table 1**, **Figs. 4**, **S1, Table S3, Files S3-S5**). Though novel *fim* alleles were
359 transient in vessels 2 and 3, in chemostat 1 a FimH Asn54Lys variant rose to a frequency of 70%

360 by generation 150, temporarily displacing high-fitness alleles in *rho*, *malK* and upstream *mglB*
361 (**File S3**). Because *fimH* mutants demonstrated an increased capacity for biofilm formation (data
362 not presented), a recurrent issue in chemostat experiments, but did not acquire any of the mutations
363 expected to enhance glucose metabolism, *fimH* mutations were likely related to chemostat
364 persistence rather than to competition for limiting substrate.
365

366 ## Discussion and Conclusions

367 ***History matters: ancestry influences evolutionary trajectory*** The tempo and trajectory of a clonal
368 population depend on its genetic point of departure. Our departure point was a founder that
369 harbored nonsense mutations in mismatch repair (MutY, Leu299*), and in housekeeping and
370 stationary phase sigma factors (RpoD, Glu26* and RpoS, Gln33*), but also carried an
371 amber/ochre/opal nonsense tRNA suppressor. Populations originating from such a founder would
372 not only have an increased mutational load but also the capacity to tolerate those mutations, in
373 particular nonsense mutations that would otherwise result in complete loss-of-function.
374 Laboratory evolution studies have borne out the idea that loss-of-function mutations can
375 be significant drivers of adaptation [20,80-82]. Metabolic network re-programming via modulation
376 of existing function can occur much faster than the evolution of new pathways via mutation [81],
377 and in many cases nonsense mutations or deletions confer greater fitness benefit than missense
378 mutations affecting the same gene [82]. However, loss of function often comes at the expense of
379 metabolic flexibility, limiting the ability of evolved clones to compete in alternative environments
380 [20]. RpoS has been shown to be a high-value target of selection under nutrient limitation: under
381 low-nutrient conditions RpoS normally outcompetes RpoD for binding to RNA polymerase,
382 repressing genes required for growth and cell division and activating those required to enter

383 stationary phase [78,83]. *rpoS* mutants thus continue to divide under conditions where wild-type
384 cells arrest. In this respect, our genetic 'point of departure' could be viewed as being pre-adapted
385 to life under glucose limitation. However, the combined phenotypic effect of ancestral *rpoS* and
386 *rpoD* nonsense mutations in a suppressor background is murky and raises the question of whether
387 this combination of mutations is favorable under glucose limitation, merely tolerated or
388 detrimental. Despite the fact that many changes we observed (*galS*, upstream *mglB*, *hfq*) enhance
389 glucose assimilation, are predictable, occur repeatedly and rise to high frequency, we also saw the
390 persistence of clones with none of these mutations that instead carry intragenic suppressors of the
391 nonsense mutation in *rpoD* (*26→Asp and *26→Gln, chemostat 3 **Fig. 4**) or a duplication that
392 includes *rpoD*. If rapid adaptation can be driven by loss-of-function but occurs at the expense of
393 metabolic flexibility, nonsense mutations have a distinct advantage over deletions in that reversion
394 or suppression is possible should environmental conditions change [20].

395 Another ancestral allele that we expected to influence evolutionary trajectories was an
396 A→T CRP binding site mutation 224 bp upstream of the acetate scavenging enzyme, *acs* (acetyl-
397 CoA synthetase). This mutation alters regulation of the *acs-pta* operon such that the ancestor
398 poorly assimilates acetate excreted during growth under continuous glucose-limitation, opening
399 up a secondary resource for novel mutants that can [34]. Here, we uncovered no evidence for the
400 type of cross-feeding described in prior reports [29,30,33]. This result was not unanticipated, as
401 evidence for cross-feeding polymorphisms was observed in only half the evolution experiments
402 founded by this ancestor or its close relatives [33]. Moreover, a recent model [31] defining the
403 boundary conditions for cross-feeding to evolve in a chemostat showed that such an outcome is
404 sensitive to variation in dilution rate as well as to the relative fitness of *de novo* mutants that gain
405 access to secondary metabolites. Subtle differences in either of these parameters may account for

406 why we saw no evidence for acetate/glycerol/formate cross-feeding in our experiments. The
407 absence of such interactions may also be due to the fact that no variants arose at loci where
408 mutations have been implicated in cross-feeding evolution: *acs* (acetyl CoA synthetase), *lpd*
409 (lipoamide dehydrogenase) and *ptsI* (phosphoenolpyruvate phosphotransferase).

410 As expected, the ancestral *acs-pta* defect resulted in appreciable levels of residual acetate
411 (~45-90 μM) at the onset of our experiments (**Fig. S3**). While we uncovered no evidence for the
412 evolution of secondary resource specialists [30,31,33] and refs therein), residual acetate levels
413 consistently fell below detection limit by generation 200. Thus, adaptive mutants arising here
414 found other ways than cross-feeding to metabolize all available carbon. One possible work-around
415 may involve the *pgi* locus, which was second only to *galS* in the total number of mutations
416 recovered (**Table S2**). Generation 200 coincides with the emergence of mutant *pgi* alleles in all
417 three populations. In chemostat experiments with *pgi* deletion mutants, Yao et al. found that in the
418 absence of Pgi, glucose uptake rate drops slightly compared to wild-type, but no overflow acetate
419 is produced and biomass yield is unchanged [84].

420 ***Population and clone sequencing open up a detailed view of the full spectrum of***
421 ***beneficial mutations and how that spectrum changes over time*** High-coverage, whole-genome,
422 whole-population sequencing makes it possible to discover every new allele reaching ≥1%
423 frequency in a population of $>10^{10}$ cells. Because alleles are highly unlikely to reach such
424 frequencies by draft, all were either transiently beneficial or hitchhiking with alleles that were.
425 This depth of analysis opens up a richly-detailed view of the spectrum of beneficial mutations
426 arising in *E. coli* under constant resource limitation. Periodic whole-population sequencing allows
427 patterns to be discerned as to how these spectra change over time, while clone sequencing makes
428 it possible to establish linkage relations among novel alleles and represent their collective fate as

429 evolving lineages. Multiple patterns emerge from these analyses. *First,* new alleles accumulate in
430 replicate populations at similar rates, and the proportion of alleles that are missense, nonsense,
431 synonymous, or noncoding remains fairly constant; the great majority is either missense (60-70%)
432 or nonsense (5-10%). *Second,* the distribution of new mutations across the genome is skewed, with
433 only a few dozen of the more than 1,000 mutated genes having a significant number of mutations;
434 yet even among these most frequently mutated genes, few *de novo* mutations fix. *Third,* by clonal
435 sequencing we are able to determine that many, independent linages co-exist and compete within
436 the culture. Thus, evolutionary dynamics in these populations is governed by clonal interference
437 rather than by clonal replacement or reinforcement.

438 A *fourth* pattern to emerge is widespread parallelism in regulatory evolution. Both across
439 and within populations, the same genes are mutated again and again, often at exactly the same
440 nucleotide position in independent replicates, and sometimes in independent lineages co-evolving
441 in the same vessel. Many of these genes (*galS*, *malT*, *malK*, upstream *mglB*, *hfq*, *rho*) act in
442 processes related to the transport and assimilation of the limiting nutrient, glucose. However, in
443 most cases the mutations recovered alter regulation of these processes, and not the structural
444 proteins that carry them out.

445 A *fifth* pattern relates to the order of beneficial mutations and the influence that order has
446 on dynamics. Consistent with previous reports, mutations that increase glucose flux across the
447 inner membrane (*galS*, upstream *mglB*) occur early and precede those that increase flux across the
448 outer membrane (*malK*/*malT*, *hfq*, *rho*). In both cases, mutations in binding partners
449 (GalS/upstream *mglB* and MalT/MalK) rarely occur in the same clone, and the order in which they
450 occur can lead to either a sweep (upstream *mglB* clones quickly displace *galS* clones) or clonal
451 interference (*malT* and *malK* clones can co-exist). Other alleles appear to emerge later in the

452 experiment and nearly always together: clones with existing mutations in the *mal* operon acquire
453 subsequent mutations in *hfq* and *opgH*, regardless of which gene is altered first and which alleles
454 are already present in the population **(Fig. S1)**. These patterns are echoed by the genotypes reported
455 by Kinnersley et al. [22] in which glucose scavenger CV103 has mutations in *malK*, *opgG* and *hfq*
456 while acetate specialist CV101 only carries a mutation in *malT*.

457 Similar experiments carried out by Maharjan et al. [32] using *E. coli* BW2952
458 demonstrated that under continuous glucose limitation population-level phenotypic changes are
459 often the result of multiple soft sweeps by combinations of beneficial mutations. While we did not
460 assay clone phenotypes, multiple alleles of *galS*, *hfq* and *opgH* appear to sweep our populations in
461 concert suggesting a similar pattern in which a phenotypic effect (reduced expression of a
462 particular gene) is favored, but has different genetic bases in co-existing lineages. At the clone
463 level, BW2952 also exhibits sign epistasis between mutations in *rpoS*/*hfq* and *galS*/*malT* [32,52].
464 In our experiments, we did not uncover evidence of sign epistasis between the ancestral *rpoS* allele
465 and *hfq*: by generation 250, over 50% of clones in populations 1 and 3 carry mutations in both
466 genes. Maharjan et al (2013) proposed that fitness deficits exhibited by *rpoS*/*hfq* double mutants
467 may be the result of altered cell division [62,85] and that *hfq* mutations enhance glucose uptake
468 during slow growth, but diminish viability when cells are dividing rapidly. Hfq deletion mutants
469 exhibit cell division anomalies due to elevated expression of cell division proteins, including FtsZ
470 [86,87]. Recent work by Hill et al. (2013) has shown that during fast growth OpgH (which in our
471 experiments is nearly always mutated alongside *hfq*) binds FtsZ to postpone cell division. Thus, it
472 may be that negative fitness effects experienced by *hfq*-*rpoS* double mutants are the result of cell
473 division anomalies mitigated by mutations in *opgH*. It is noteworthy in this regard that whereas
474 cells in the Maharjan et al. experiments experienced a dilution rate of D=0.1 hr^{-1}, those in

evolutions performed by Adams, Helling and colleagues were doubling twice as fast (D=0.2 h^{-1}). Thus, this discrepancy may be a manifestation of trade-offs between glucose uptake and cell viability.

Finally, some mutations occur repeatedly and are likely beneficial, but their dynamics are unpredictable (e.g. beneficial mutations in transcriptional terminator *rho* sweep when they co-occur with beneficial mutations upstream of *mglB,* but otherwise remain at low frequency (**Table S3, Fig. S1**). This dependence on genetic context, or "quasi-hitchhiking", of beneficial mutations was previously observed by Lang et al. (2013) in evolving yeast populations and may be consistent feature of microbial evolution experiments that becomes observable when populations are sequenced at high depth of coverage and sufficient temporal resolution [14].

The evolution of population genetic complexity. Szostak, Hazen and others [88,89] argue that a biological system's complexity should be evaluated in terms of its functional information content. Although the total number of alleles in an evolving population at any given time-point is information content, it is functional only in how it is integrated among the lineages co-existing at that time-point. Our approach of integrating population sequencing with clone sequencing makes it possible to estimate the pace and extent with which complexity, measured as lineage-specific functional information [88,89], emerges in replicate evolving populations originating from a common ancestor. Implicit in our perspective is the assumption that the sequence differences by which lineages can be distinguished have physiological and fitness consequences. For each population, we calculated at 50-generation intervals three measures of complexity: Shannon's Entropy (H), equitability (H/H_{max}) and normalized population richness (Lineage counts) (**Fig. S5**). All three measures increased during the course of evolution, but with a different tempo in each population. Complexity increased in population 3 with no indication of reaching an asymptote by

generation 300 when the experiment was terminated. Populations 1 and 2 reached asymptotes after ~400 generations, following a steady increase in population 1, but a nearly-300 generation period of stasis in Population 2. While longer experiments are clearly called for, our finding that longer-term experiments appear to reach an asymptote in complexity is consistent with theoretical [90] and empirical [91]observations that fitness plateaus when microbes are cultured by serial transfer or in chemostats, even starting with mutator strains [92], and that complexity itself may plateau when its evolution is simulated using RNA-like replicators [93]. It is intriguing to contemplate the possibility that there may be a limit to the level of clonal interference that can be sustained in asexual populations once all avenues for large fitness gains have been exhausted. Indeed, it was recently shown using lineage tracking, that while fixation of an adaptive mutant causes a stochastic crash in diversity, the generation of new adaptive mutants within such a fixing lineage is expected to generate new diversity, such that a longer term steady state level of diversity will be achieved [94].

Previous evolution experiments founded with the ancestor used here often resulted in stable sub-populations supported by cross-feeding. In the present experiments neither the spectrum of observed mutations nor the structure of clone phylogenies provides evidence for the evolution of this type of trophic interaction. Instead, the possibility of a plateau in complexity, coupled with the finding that every population has driven residual metabolites close to their detection limit, suggest that these populations converge on an adaptive peak by diverse mechanisms but that clonal interference keeps adaptive lineages off the summit, confined to exploring the many roads by which the summit can be approached.

Materials and Methods

521 ***Strains, media and culture conditions.*** *Escherichia coli* JA122, population samples and clones
522 were maintained as permanent frozen stocks and stored at -80°C in 20% glycerol. Davis minimal
523 medium was used for all liquid cultures with 0.025% glucose added for batch cultures and 0.0125%
524 for chemostats, as previously described [34]. Chemostat cultures were initiated using colonies
525 picked from Tryptone Agar (TA) plate inoculated with JA122, and outgrown in Davis minimal
526 batch medium overnight. Chemostats were maintained at 30°C with a dilution rate of ≈ 0.2 hr^{-1} for
527 300-500 generations. Every other day culture density was assessed by measuring absorbance
528 spectrophotometrically at A_{550}. Every other day, population samples were archived at -80°C, and
529 assayed for purity by plating serial dilutions on TA and examining Colony Forming Units (CFU)
530 that arose following 24-hr incubation at 30°C. When necessary, chemostats were re-started from
531 frozen stocks (chemostat 1: generation 217; chemostat 2: generation 410; chemostat 3 generation
532 251). At each sequencing time-point, 50 mL of culture was pelleted then stored at -80°C for DNA
533 extraction. For clone sequencing, entire colonies were picked from TA plates inoculated from
534 glycerol stocks, and re-archived in 96-well plate format.
535 ***Metabolite assays.*** 10 mL of sterile, cell-free chemostat filtrate was concentrated 20-fold by
536 lyophilization (Labconco 4.5 Liter Freeze Dry System), then re-suspended in 0.5 mL sterile
537 Millipore water. Residual glucose and residual acetate concentrations were determined on
538 concentrated filtrate. Glucose was assayed enzymatically using the High Sensitivity Glucose
539 Assay Kit (Sigma-Aldrich, Cat# MAK181), while acetate concentration was determined using the
540 Acetate Colorimetric Assay Kit (Sigma-Aldrich, Cat# MAK086). Results presented in **Fig. S3**
541 represent means ± SEM of duplicate assays.
542 ***Population sequencing.*** Bacterial DNA was prepared using the DNeasy Blood and Tissue Kit
543 (Qiagen, cat. 69504) following the manufacturer's guidelines. For population sequencing, 5 x 10^{10}

544 cells, collected from every 50 generations in three chemostat vessels (up to 500 generations in
545 vessels 1 and 2, and up to 300 generations in vessel 3, 29 samples total) and frozen as pellets,
546 yielded 10-20µg of DNA. Following Proteinase K treatment, RNaseA treatment was used (20µL
547 10mg/mL RNAse A, 2 min at room temperature) to avoid degraded RNA from visually obscuring
548 size selection during library preparation. Samples were split into two columns to avoid
549 overloading. Bacterial DNA was sheared to a 150-200bp fragment size using a Covaris S2 series
550 sonicator (6min, Duty=5%, Intensity=3, Cycles/Burst=200), and was then ligated to barcoded
551 adapters as described [95], except that 200bp fragments were size selected after adapter ligation
552 (to maximize the fidelity of sequencing, by reading each fragment in both directions). Six barcoded
553 libraries were combined and sequenced on each lane of HiSeq 2000 Illumina Sequencer.
554 ***Variant calling from population sequencing with CLC Genomics Workbench 7.5*** Illumina reads
555 were trimmed (removing adapters on both ends) and stringently mapped (Mismatch cost 2,
556 Insertion cost 3, Deletion cost 3, Length fraction 1.0, Similarity fraction 0.97) to the reference
557 sequence (WIS_MG1655_m56). Variants were called with the following parameters: minimum
558 frequency 1%, minimal coverage 100, minimum count 2, and base quality filtering (neighborhood
559 radius 5, minimum central quality 15 and minimum neighborhood quality 20). Sequencing data
560 uncovered low-level contamination of whole population samples with *Serratia liquifaciensis.* We
561 therefore first determined proportion of contaminating reads by mapping population sequencing
562 to *S. liquifaciensis* genome and then removed SNPs with frequency closely tracking percentage of
563 contamination (between 1 and 5%) that matched *S. liquifaciensis* sequence.
564 ***Selection of clones for sequencing.*** Allele frequencies for each chemostat were examined at each
565 time point, and the time-point at which there was the largest number of alleles present at 5% or
566 greater frequency was chosen for the isolation of clones for whole genome sequencing. The

567 rationale for this was that it would afford us the greatest opportunity to phase as many high
568 frequency alleles as possible.

569 ***Clonal DNA preparation.*** A colony was re-suspended in 300µL of sterile ddH_20 with 17%
570 glycerol and stored in three aliquots at -80°C. 100 µL of glycerol stock were used for DNA
571 preparation. After removing glycerol (using MultiScreen High Volume Filter Plates with 0.45 µm
572 Durapore membrane, Millipore MVHVN4525), cell were resuspended in 500µL LB and grown
573 overnight at 30°C without shaking in deep well plates. Cells were collected again using filter plates
574 and subjected to DNeasy 96 Blood and Tissue Kit (Qiagen 69581) (yielding 4-15µg per strain).

575 ***Clonal libraries preparation and sequencing.*** Multiplexed sequencing libraries from clones were
576 prepared using the Nextera kit (Illumina catalog # FC-121-1031 and # FC-121-1012) as described
577 in [96], starting with 1-4ng of genomic DNA. Resulting libraries from each 96-well plate were
578 pooled at an equal volume. Resulting pooled libraries were analyzed on the Qubit and Bioanalyzer
579 platforms and sequenced on HiSeq 2000 (one lane per 96 clone pool). All raw sequencing data are
580 available from the SRA under BioProject ID PRJNA517527.

581 ***Variant calling from clonal sequencing with CLC Genomics Workbench 7.5*** Short reads with
582 adapters removed were mapped to the reference with the same parameters as above, except the
583 length fraction was set to 0.5, and the similarity fraction to 0.8. Variants were called with a
584 minimum frequency 80%, minimum count 2 and the same base quality filtering as above.

585 ***Generation of phylogenies.*** For each chemostat, SNP and indel events for all 96 clones and the
586 ancestor JA122 were concatenated and re-coded as binary characters (i.e. presence/absence with
587 the ancestral state composed of all zeroes) and assembled into character matrices. PAUP ver.
588 4.0a149 was used to generate Camin-Sokal parsimony trees using the ancestor as the outgroup
589 under the assumption that reversions were extremely unlikely due to the extreme transversion bias

590 [97,98]. Tree files (.tre) were loaded into the Interactive Tree of Life (iTOL) web service for
591 character mapping and figure generation [99].

592 ***Determining genes with an excess of mutations***. To identify genes with an excess of mutations,
593 we first determined the overall density of mutations as:

594 $\rho = M/L$, where M is the total number of mutations, and L is the length of the genome.

595 The probability of a given mutation landing in a segment of length l, is:

596 $\lambda = \rho \times l$

597 To calculate the p-value of n mutations landing in a segment of length l, we assume a Poisson
598 sampling process of a mutation landing in a given segment, and thus use:

599 $$p = \sum_{i=n}^{\infty} \frac{\lambda^i \times e^{-\lambda}}{i!}$$

600 though in practice, we capped i arbitrarily at 50, as continually summing at i >50 does not
601 appreciably affect the calculated p-value. For a given segment, we calculated the number of
602 segments that would be expected to have p-value as good or better, as the number of tested
603 segments multiplied by the p-value. From this, we also determined a false positive rate.

604 ***Generation of Muller diagrams***. Based on both the clonal sequencing we were able to determine
605 which mutations were in which lineages together, and from both the clonal and population
606 sequencing an approximate order of those mutations (though this was not exhaustive for all
607 mutations). Using these data, we developed a lineage file format that described which mutations
608 occurred in which lineages, and which lineages descended from one another, and used a custom
609 Perl script that combined this information with the allele frequencies over time from the population
610 sequencing to generate a graphical representation of the evolutionary dynamics, often referred to
611 as a Muller diagram.

Acknowledgements

The authors thank Matthew Herron and Eugene Kroll for their careful reading of the manuscript and their thoughtful suggestions for its improvement.

Author Contributions

Conceived and designed the experiments: GS MK FR. Performed the experiments: MK and KS. Analyzed the data: GS KS MK JB DY FR. Contributed reagents/materials/analysis tools: GS. Wrote the paper: MK GS FR

621 ## Main Figure legends

622 **Fig. 1. Input of de novo mutations.** The rate at which novel alleles appear, and the proportion of
623 synonymous, missense and nonsense mutations, and non-coding mutations are consistent across
624 all three replicate evolutions.

625

626 **Fig. 2. Most *de novo* mutations only reach low allele frequencies, and experience pervasive**
627 **clonal interference.** (A) Histogram of maximum allele frequencies from three replicate
628 evolutions, (B) Final versus maximum allele frequency for each *de novo* mutation, shows most
629 mutations are at a lower frequency at the end of the experiment than they were at their maximum.

630 **Fig. 3. Isolated clones are representative of the populations from which they are drawn.**
631 Mutation frequencies for population and clonal sequencing for mutations identified in both
632 datasets at the same time-point shows similar frequencies.
633

634 **Fig. 4. Clone phylogenies.** Phylogenies depicting relationships among sequenced clones isolated
635 from chemostats when allelic diversity attained its maximum. Distributions of different *malK*,
636 *malT*, *fimH, hfq* and *opgH* alleles are indicated by colored bars. For each gene, all alleles observed
637 in the dataset are numbered (see File S2 for details of which number corresponds to which allele
638 for each gene). Underlined numbers denote alleles independently observed in more than one
639 chemostat, while numbers marked with an asterisk appear to have arisen more than once within
640 the same vessel. Grey shading delineates clades comprised of clones that have not acquired the
641 standard mutations related to enhanced glucose uptake and instead carry variant *fimH* alleles that
642 contribute to biofilm formation. Bracketed clones in chemostat 3 exhibited mutations expected to
643 revert the ancestral nonsense mutations in the housekeeping gene encoding sigma factor RpoD.

644 **Fig. 5**. **Muller diagrams.** Evolutionary dynamics of adaptive lineages, deduced from combining
645 whole-population whole-genome sequence data and whole-genome sequence data of individual
646 clones isolated from each chemostat at the time-point where allelic diversity reached its maximum
647 value. Select genes are indicated in the plots – see Figure S1 and Supplementary Files S3-S5 for
648 additional details. Also note, most mutations that went extinct by the sampling timepoint are not
649 shown. See Figure S1 for their relative frequencies.
650
651 **Fig. 6**. **Overview of pathways relating some of the most frequently mutated genes to glucose**
652 **transport and metabolism.** Numbers in parentheses next to protein/gene names denote the
653 number of mutant alleles found in each chemostat population over the course of 300-500
654 generations (also see Table S3).
655
656

657 ## Supplementary Figure Legends

658 **Fig. S1. Population-level dynamics of mutations in 10 frequently hit genes show consistent**
659 **patterns.** For each panel, elapsed generations are depicted on the x-axis and the height of each
660 grey box represents a frequency of 100%. Cumulative frequencies for all alleles of a given gene
661 present in the population at each time point were calculated and are represented as colored plots.
662 (A) chemostat 1 (B) chemostat 2 (C) chemostat 3.
663
664 **Fig. S2. MalK/MalT population dynamics.** Mutant alleles of both LamB regulators, *malT* and
665 *malK*, seldom co-occur in the same lineage, and when they do, those lineages fail to go to high
666 frequency.
667
668 **Fig. S3**. **Cell density and residual metabolite concentrations**. Chemostat populations at steady
669 state exhibit balanced growth, where population size remains constant and the limiting substrate,
670 glucose is near its detection limit. As expected, populations initially produce the overflow
671 metabolite acetate, as the founder carries a mutation that dysregulates acetyl CoA synthetase, the
672 chief route by which *E. coli* assimilates low levels of this metabolite.
673
674 **Fig. S4. Mutations in glucosyltransferase *opgH* occur repeatedly and go to high frequency.**
675
676 **Fig. S5. Patterns of change in population genetic complexity.** Shannon's Entropy [H],
677 Equitability [H/Hmax] and normalized population Richness [Lineage counts] were calculated at
678 50 generation intervals for each of three replicate evolution experiments. Shannon's entropy is an
679 effective metric of population diversity as it accounts for both lineage richness (the number of

lineages observed) and the relative abundance of different lineages (evenness). Lineage richness was normalized between zero and one by dividing the number of observed lineages, *S*, by the maximum *S* observed over the course of each experiment.

Main Tables

Table 1. Characteristics of frequently mutated genes. Each asterisk indicates an allele that arose more than once independently, either within or between vessels.

Supplementary Tables

Table S1. Key mutations that distinguish ancestral strain JA122 from K12 (MG1655)

Table S2. Beneficial alleles

Table S3. Population allele frequencies for frequently mutated genes

Table S4. Identical mutations arise within and among replicate evolution experiments.

Table S5. Fixed alleles among replicate populations ("fixed" defined as >98% at any time point between generation 50 and 500).

Supplementary Data Files

File S1. Identity and frequencies of mutations detected via population sequencing.

File S2. Alleles mapped onto clone phylogenies represented in main Fig. 4.

File S3. Muller diagrams for novel alleles arising in chemostat 1, showing details for each lineage

File S4. Muller diagrams for novel alleles arising in chemostat 2, showing details for each lineage

File S5. Muller diagrams for novel alleles arising in chemostat 3, showing details for each lineage.

703 **Table 1 Frequently mutated genes.** Each asterisk indicates an allele that
704 arose more than once independently, either within or between vessels.

Rank	Gene	Observed Mutations	Expected Mutations	P-Value	FDR
Population sequencing					
1	galS***	38	0.78	6.55E-50	4.42E-45
2	hfq********	24	0.23	6.91E-40	2.33E-35
3	pgi********	35	1.23	4.54E-38	1.02E-33
4	opgH**	31	1.90	8.74E-27	1.48E-22
5	malT*********	30	2.02	8.10E-25	1.09E-20
6	malK*******	22	0.83	7.47E-24	8.40E-20
7	upstream mglB**	7	0.21	2.91E-09	2.81E-05
8	rho**	11	0.94	5.49E-09	4.64E-05
9	upstream dnaG	5	0.08	3.06E-08	2.30E-04
10	fimH***	9	0.68	4.38E-08	2.96E-04
Clonal sequencing					
1	hfq*******	26	0.1020	3.79E-53	3.68E-48
2	pgi****	17	0.5448	5.52E-20	2.68E-15
3	opgH***	17	0.8400	6.57E-17	2.13E-12
4	upstream mglB**	8	0.0925	1.22E-13	2.96E-09
5	fimH****	10	0.2982	1.17E-12	2.26E-08
6	ompR***	8	0.2377	2.05E-10	3.31E-06
7	upstream adhE*	6	0.1575	1.85E-08	0.000257
8	malT***	10	0.8935	3.98E-08	0.000482
9	proQ*	6	0.2308	1.72E-07	0.001858
10	pfkA**	6	0.3180	1.09E-06	0.010612

705

706

707

References

1. Lang GI, Desai MM (2014) The spectrum of adaptive mutations in experimental evolution. Genomics 104: 412-416.
2. Cvijovic I, Nguyen Ba AN, Desai MM (2018) Experimental studies of evolutionary dynamics in Microbes. Trends Genet 34: 693-703.
3. Cooper VS (2018) Experimental Evolution as a high-throughput screen for genetic adaptations. mSphere 3.
4. Garay E, Campos SE, Gonzalez de la Cruz J, Gaspar AP, Jinich A, et al. (2014) High-resolution profiling of stationary-phase survival reveals yeast longevity factors and their genetic interactions. PLoS Genet 10: e1004168.
5. Paradis-Bleau C, Kritikos G, Orlova K, Typas A, Bernhardt TG (2014) A genome-wide screen for bacterial envelope biogenesis mutants identifies a novel factor involved in cell wall precursor metabolism. PLoS Genet 10: e1004056.
6. Novick A, Szilard L (1950) Experiments with the chemostat on spontaneous mutations of bacteria. Proc Natl Acad Sci U S A 36: 708-719.
7. Novick A, Szilard L (1951) Experiments on spontaneous and chemically induced mutations of bacteria growing in the Chemostat. Cold Spring Harb Symp Quant Biol 16: 337-343.
8. Atwood KC, Schneider LK, Ryan FJ (1951) Periodic selection in Escherichia coli. Proc Natl Acad Sci U S A 37: 146-155.
9. Atwood KC, Schneider LK, Ryan FJ (1951) Selective mechanisms in bacteria. Cold Spring Harb Symp Quant Biol 16: 345-355.
10. Haldane JBS (1927) A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. Mathematical Proceedings of the Cambridge Philosophical Society 23: 838-844.
11. Fisher RAS (1930) The genetical theory of natural selection. Oxford: Clarendon Press.
12. Muller HJ (1932) Some genetic aspects of sex. The American Naturalist 66: 118-138.
13. Gause GF (1934) Experimental analysis of Vito Volterra's mathematical theory of the struggle for existence. Science 79: 16-17.
14. Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, et al. (2013) Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature 500: 571-574.
15. Marad DA, Buskirk SW, Lang GI (2018) Altered access to beneficial mutations slows adaptation and biases fixed mutations in diploids. Nat Ecol Evol 2: 882-889.
16. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, et al. (2015) Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature 519: 181-186.
17. Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM (2017) The dynamics of molecular evolution over 60,000 generations. Nature 551: 45-50.
18. Lauer S, Avecilla G, Spealman P, Sethia G, Brandt N, et al. (2018) Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. PLoS Biol 16: e3000069.
19. Kao KC, Sherlock G (2008) Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. Nat Genet 40: 1499-1504.
20. Kvitek DJ, Sherlock G (2013) Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. PLoS Genet 9: e1003972.

754 21. Behringer MG, Choi BI, Miller SF, Doak TG, Karty JA, et al. (2018) Escherichia coli
755 cultures maintain stable subpopulation structure during long-term evolution. Proc Natl
756 Acad Sci U S A 115: E4642-E4650.
757 22. Rozen DE, Lenski RE (2000) Long-term experimental evolution in Escherichia coli. VIII.
758 Dynamics of a balanced polymorphism. Am Nat 155: 24-35.
759 23. Herron MD, Doebeli M (2013) Parallel evolutionary dynamics of adaptive diversification in
760 Escherichia coli. PLoS Biol 11: e1001490.
761 24. Maddamsetti R, Lenski RE, Barrick JE (2015) Adaptation, clonal interference, and
762 frequency-dependent jnteractions in a long-term evolution experiment with Escherichia
763 coli. Genetics 200: 619-631.
764 25. Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome
765 regions. PLoS Genet 2: e64.
766 26. Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. Nature
767 394: 69-72.
768 27. Rozen DE, Schneider D, Lenski RE (2005) Long-term experimental evolution in Escherichia
769 coli. XIII. Phylogenetic history of a balanced polymorphism. J Mol Evol 61: 171-180.
770 28. Kinnersley M, Wenger J, Kroll E, Adams J, Sherlock G, et al. (2014) Ex uno plures: clonal
771 reinforcement drives evolution of a simple microbial community. PLoS Genet 10:
772 e1004430.
773 29. Rosenzweig RF, Sharp RR, Treves DS, Adams J (1994) Microbial evolution in a simple
774 unstructured environment: genetic differentiation in Escherichia coli. Genetics 137: 903-
775 917.
776 30. Helling RB, Vargas CN, Adams J (1987) Evolution of Escherichia coli during growth in a
777 constant environment. Genetics 116: 349-358.
778 31. Gudelj I, Kinnersley M, Rashkov P, Schmidt K, Rosenzweig F (2016) Stability of cross-
779 feeding polymorphisms in microbial communities. PLoS Comput Biol 12: e1005269.
780 32. Maharjan RP, Liu B, Feng L, Ferenci T, Wang L (2015) Simple phenotypic sweeps hide
781 complex genetic changes in populations. Genome Biol Evol 7: 531-544.
782 33. Treves DS, Manning S, Adams J (1998) Repeated evolution of an acetate-crossfeeding
783 polymorphism in long-term populations of Escherichia coli. Mol Biol Evol 15: 789-797.
784 34. Kinnersley MA, Holben WE, Rosenzweig F (2009) E Unibus Plurum: genomic analysis of
785 an experimentally evolved polymorphism in Escherichia coli. PLoS Genet 5: e1000713.
786 35. Atlung T, Nielsen HV, Hansen FG (2002) Characterisation of the allelic variation in the rpoS
787 gene in thirteen K12 and six other non-pathogenic Escherichia coli strains. Mol Genet
788 Genomics 266: 873-881.
789 36. Singaravelan B, Roshini BR, Munavar MH (2010) Evidence that the supE44 mutation of
790 Escherichia coli is an amber suppressor allele of glnX and that it also suppresses ochre
791 and opal nonsense mutations. J Bacteriol 192: 6039-6044.
792 37. Au KG, Clark S, Miller JH, Modrich P (1989) Escherichia coli mutY gene encodes an
793 adenine glycosylase active on G-A mispairs. Proc Natl Acad Sci U S A 86: 8877-8881.
794 38. Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, et al. (1997) The complete
795 genome sequence of Escherichia coli K-12. Science 277: 1453-1462.
796 39. Lee H, Popodi E, Tang H, Foster PL (2012) Rate and molecular spectrum of spontaneous
797 mutations in the bacterium Escherichia coli as determined by whole-genome sequencing.
798 Proc Natl Acad Sci U S A 109: E2774-2783.

40. de Visser JA, Rozen DE (2006) Clonal interference and the periodic selection of new beneficial mutations in Escherichia coli. Genetics 172: 2093-2100.
41. Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. Genetics 169: 2335-2352.
42. Pennings PS, Hermisson J (2006) Soft sweeps II--molecular population genetics of adaptation from recurrent mutation or migration. Mol Biol Evol 23: 1076-1084.
43. Pennings PS, Hermisson J (2006) Soft sweeps III: the signature of positive selection from recurrent mutation. PLoS Genet 2: e186.
44. Desai MM, Walczak AM, Fisher DS (2013) Genetic diversity and the structure of genealogies in rapidly adapting populations. Genetics 193: 565-585.
45. Notley-McRobb L, Ferenci T (1999) The generation of multiple co-existing mal-regulatory mutations through polygenic evolution in glucose-limited populations of Escherichia coli. Environ Microbiol 1: 45-52.
46. Hermisson J, Pennings PS (2017) Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. Methods in Ecology and Evolution 8: 700-716.
47. Jensen JD (2014) On the unfounded enthusiasm for soft selective sweeps. Nat Commun 5: 5281.
48. Fogle CA, Nagle JL, Desai MM (2008) Clonal interference, multiple mutations and adaptation in large asexual populations. Genetics 180: 2163-2173.
49. Buskirk SW, Peace RE, Lang GI (2017) Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. Proc Natl Acad Sci U S A 10.1073/pnas.1702314114.
50. Diaz Caballero J, Clark ST, Coburn B, Zhang Y, Wang PW, et al. (2015) Selective Sweeps and Parallel Pathoadaptation Drive Pseudomonas aeruginosa Evolution in the Cystic Fibrosis Lung. MBio 6: e00981-00915.
51. Ferenci T (2001) Hungry bacteria--definition and properties of a nutritional state. Environ Microbiol 3: 605-611.
52. Maharjan RP, Ferenci T (2013) Epistatic interactions determine the mutational pathways and coexistence of lineages in clonal Escherichia coli populations. Evolution 67: 2762-2768.
53. Manch K, Notley-McRobb L, Ferenci T (1999) Mutational adaptation of Escherichia coli to glucose limitation involves distinct evolutionary pathways in aerobic and oxygen-limited environments. Genetics 153: 5-12.
54. Notley-McRobb L, Ferenci T (1999) Adaptive mgl-regulatory mutations and genetic diversity evolving in glucose-limited Escherichia coli populations. Environ Microbiol 1: 33-43.
55. Notley-McRobb L, Ferenci T (2000) Experimental analysis of molecular events during mutational periodic selections in bacterial evolution. Genetics 156: 1493-1501.
56. Notley-McRobb L, Seeto S, Ferenci T (2003) The influence of cellular physiology on the initiation of mutational pathways in Escherichia coli populations. Proc Biol Sci 270: 843-848.
57. Ferenci T (1996) Adaptation to life at micromolar nutrient levels: the regulation of Escherichia coli glucose transport by endoinduction and cAMP. FEMS Microbiol Rev 18: 301-317.
58. Death A, Ferenci T (1993) The importance of the binding-protein-dependent Mgl system to the transport of glucose in Escherichia coli growing on low sugar concentrations. Res Microbiol 144: 529-537.

845 59. Death A, Notley L, Ferenci T (1993) Derepression of LamB protein facilitates outer
846 membrane permeation of carbohydrates into Escherichia coli under conditions of nutrient
847 stress. J Bacteriol 175: 1475-1483.
848 60. Tweeddale H, Notley-McRobb L, Ferenci T (1999) Assessing the effect of reactive oxygen
849 species on Escherichia coli using a metabolome approach. Redox Rep 4: 237-241.
850 61. Geanacopoulos M, Adhya S (1997) Functional characterization of roles of GalR and GalS as
851 regulators of the gal regulon. J Bacteriol 179: 228-234.
852 62. Maharjan R, McKenzie C, Yeung A, Ferenci T (2013) The basis of antagonistic pleiotropy in
853 hfq mutations that have opposite effects on fitness at slow and fast growth rates. Heredity
854 (Edinb) 110: 10-18.
855 63. Maharjan R, Zhou Z, Ren Y, Li Y, Gaffe J, et al. (2010) Genomic identification of a novel
856 mutation in hfq that provides multiple benefits in evolving glucose-limited populations of
857 Escherichia coli. J Bacteriol 192: 4517-4521.
858 64. Richet E, Joly N, Danot O (2005) Two domains of MalT, the activator of the Escherichia coli
859 maltose regulon, bear determinants essential for anti-activation by MalK. J Mol Biol 347:
860 1-10.
861 65. Dardonville B, Raibaud O (1990) Characterization of malT mutants that constitutively
862 activate the maltose regulon of Escherichia coli. J Bacteriol 172: 1846-1852.
863 66. Bohm A, Diez J, Diederichs K, Welte W, Boos W (2002) Structural model of MalK, the
864 ABC subunit of the maltose transporter of Escherichia coli: implications for mal gene
865 regulation, inducer exclusion, and subunit assembly. J Biol Chem 277: 3708-3717.
866 67. Moller P, Overloper A, Forstner KU, Wen TN, Sharma CM, et al. (2014) Profound impact of
867 Hfq on nutrient acquisition, metabolism and motility in the plant pathogen
868 Agrobacterium tumefaciens. PLoS One 9: e110427.
869 68. Maharjan RP, Ferenci T, Reeves PR, Li Y, Liu B, et al. (2012) The multiplicity of divergence
870 mechanisms in a single evolving population. Genome Biol 13: R41.
871 69. Cardinale CJ, Washburn RS, Tadigotla VR, Brown LM, Gottesman ME, et al. (2008)
872 Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli.
873 Science 320: 935-938.
874 70. Banerjee S, Chalissery J, Bandey I, Sen R (2006) Rho-dependent transcription termination:
875 more questions than answers. J Microbiol 44: 11-22.
876 71. Peters JM, Mooney RA, Kuan PF, Rowland JL, Keles S, et al. (2009) Rho directs widespread
877 termination of intragenic and stable RNA transcription. Proc Natl Acad Sci U S A 106:
878 15406-15411.
879 72. Ciampi MS (2006) Rho-dependent terminators and transcription termination. Microbiology
880 152: 2515-2528.
881 73. Colonna B, Hofnung M (1981) rho Mutations restore lamB expression in E. coli K12 strains
882 with an inactive malB region. Mol Gen Genet 184: 479-483.
883 74. Nakashima N, Ohno S, Yoshikawa K, Shimizu H, Tamura T (2014) A vector library for
884 silencing central carbon metabolism genes with antisense RNAs in Escherichia coli. Appl
885 Environ Microbiol 80: 564-573.
886 75. Hill NS, Buske PJ, Shi Y, Levin PA (2013) A moonlighting enzyme links Escherichia coli
887 cell size with central metabolism. PLoS Genet 9: e1003663.
888 76. Lange R, Hengge-Aronis R (1994) The cellular concentration of the sigma S subunit of RNA
889 polymerase in Escherichia coli is controlled at the levels of transcription, translation, and
890 protein stability. Genes Dev 8: 1600-1612.

891 77. King T, Ishihama A, Kori A, Ferenci T (2004) A regulatory trade-off as a source of strain
892 variation in the species Escherichia coli. J Bacteriol 186: 5614-5620.
893 78. Notley-McRobb L, King T, Ferenci T (2002) rpoS mutations and loss of general stress
894 resistance in Escherichia coli populations as a consequence of conflict between
895 competing stress responses. J Bacteriol 184: 806-811.
896 79. Versalovic J, Koeuth T, Britton R, Geszvain K, Lupski JR (1993) Conservation and
897 evolution of the rpsU-dnaG-rpoD macromolecular synthesis operon in bacteria. Mol
898 Microbiol 8: 343-355.
899 80. Hutchins PR, Miller SR (2017) Genomics of variation in nitrogen fixation activity in a
900 population of the thermophilic cyanobacterium Mastigocladus laminosus. ISME J 11: 78-
901 86.
902 81. Hottes AK, Freddolino PL, Khare A, Donnell ZN, Liu JC, et al. (2013) Bacterial adaptation
903 through loss of function. PLoS Genet 9: e1003617.
904 82. Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, et al. (2016) Development of a
905 Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast.
906 Cell 166: 1585-1596 e1522.
907 83. Phan K, Ferenci T (2013) A design-constraint trade-off underpins the diversity in
908 ecologically important traits in species Escherichia coli. ISME J 7: 2034-2043.
909 84. Yao R, Hirose Y, Sarkar D, Nakahigashi K, Ye Q, et al. (2011) Catabolic regulation analysis
910 of Escherichia coli and its crp, mlc, mgsA, pgi and ptsG mutants. Microb Cell Fact 10:
911 67.
912 85. Vecerek B, Rajkowitsch L, Sonnleitner E, Schroeder R, Blasi U (2008) The C-terminal
913 domain of Escherichia coli Hfq is required for regulation. Nucleic Acids Res 36: 133-
914 143.
915 86. Takada A, Wachi M, Nagai K (1999) Negative regulatory role of the Escherichia coli hfq
916 gene in cell division. Biochem Biophys Res Commun 266: 579-583.
917 87. Zambrano N, Guichard PP, Bi Y, Cayrol B, Marco S, et al. (2009) Involvement of HFq
918 protein in the post-transcriptional regulation of E. coli bacterial cytoskeleton and cell
919 division proteins. Cell Cycle 8: 2470-2472.
920 88. Hazen RM, Griffin PL, Carothers JM, Szostak JW (2007) Functional information and the
921 emergence of biocomplexity. Proc Natl Acad Sci U S A 104 Suppl 1: 8574-8581.
922 89. Szostak JW (2003) Functional information: Molecular messages. Nature 423: 689.
923 90. Gordo I, Campos PR (2013) Evolution of clonal populations approaching a fitness peak. Biol
924 Lett 9: 20120239.
925 91. de Visser JA, Lenski RE (2002) Long-term experimental evolution in Escherichia coli. XI.
926 Rejection of non-transitive interactions as cause of declining rate of adaptation. BMC
927 Evol Biol 2: 19.
928 92. Maharjan RP, Liu B, Li Y, Reeves PR, Wang L, et al. (2013) Mutation accumulation and
929 fitness in mutator subpopulations of Escherichia coli. Biol Lett 9: 20120961.
930 93. Takeuchi N, Hogeweg P (2008) Evolution of complexity in RNA-like replicator systems.
931 Biol Direct 3: 11.
932 94. Blundell JR, Schwartz K, Francois D, Fisher DS, Sherlock G, et al. (2018) The dynamics of
933 adaptive genetic diversity during the early stages of clonal evolution. Nat Ecol Evol
934 10.1038/s41559-018-0758-1.

95. Schwartz K, Wenger JW, Dunn B, Sherlock G (2012) APJ1 and GRE3 homologs work in concert to allow growth in xylose in a natural Saccharomyces sensu stricto hybrid yeast. Genetics 191: 621-632.
96. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM (2014) Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science 344: 1519-1522.
97. Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental Escherichia coli population. Nature 489: 513-518.
98. Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). 4 ed. Sunderland, Massachusetts: Sinauer Associates.
99. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23: 127-128.

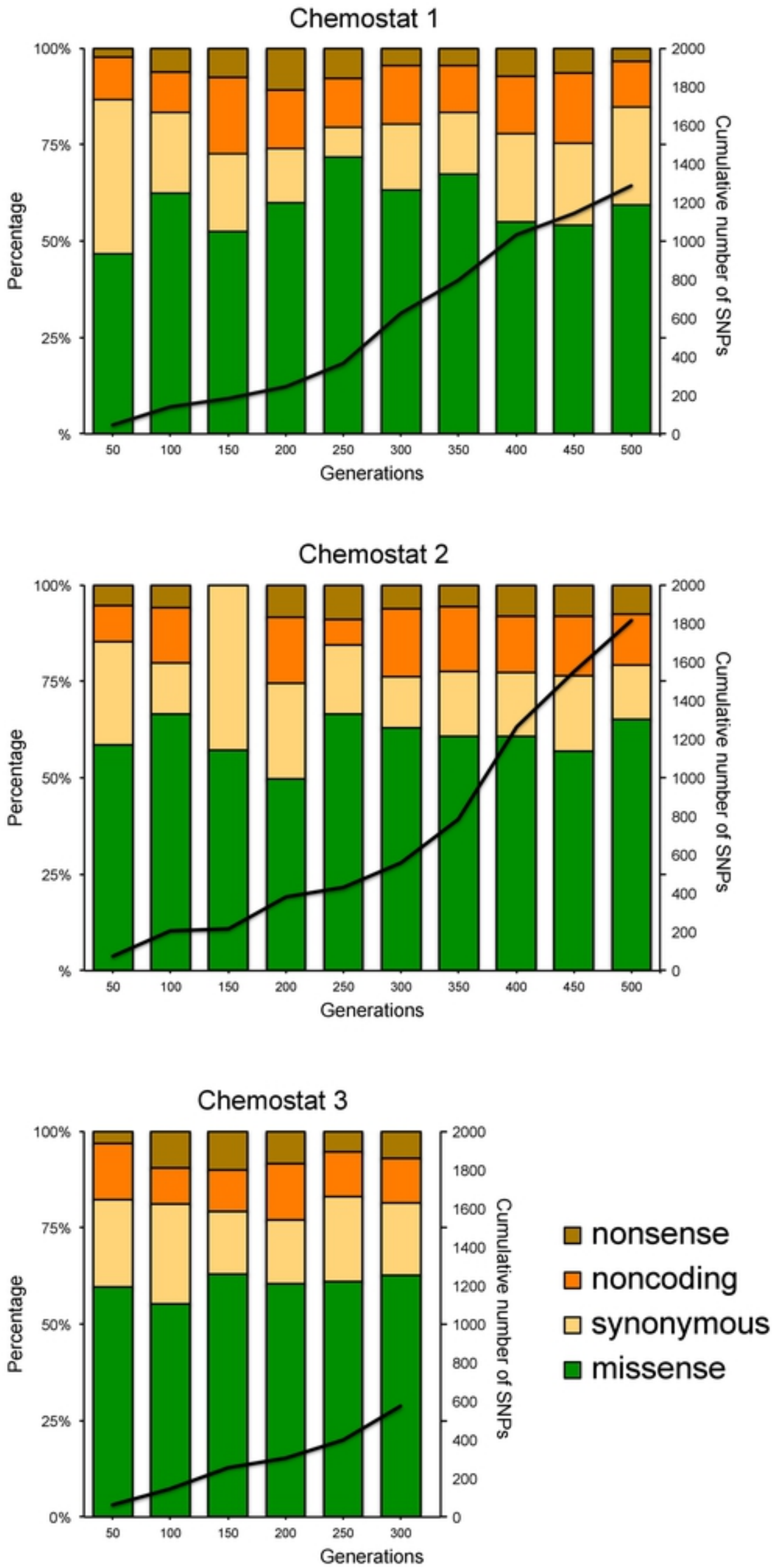

Fig. 1 Input of de novo mutations

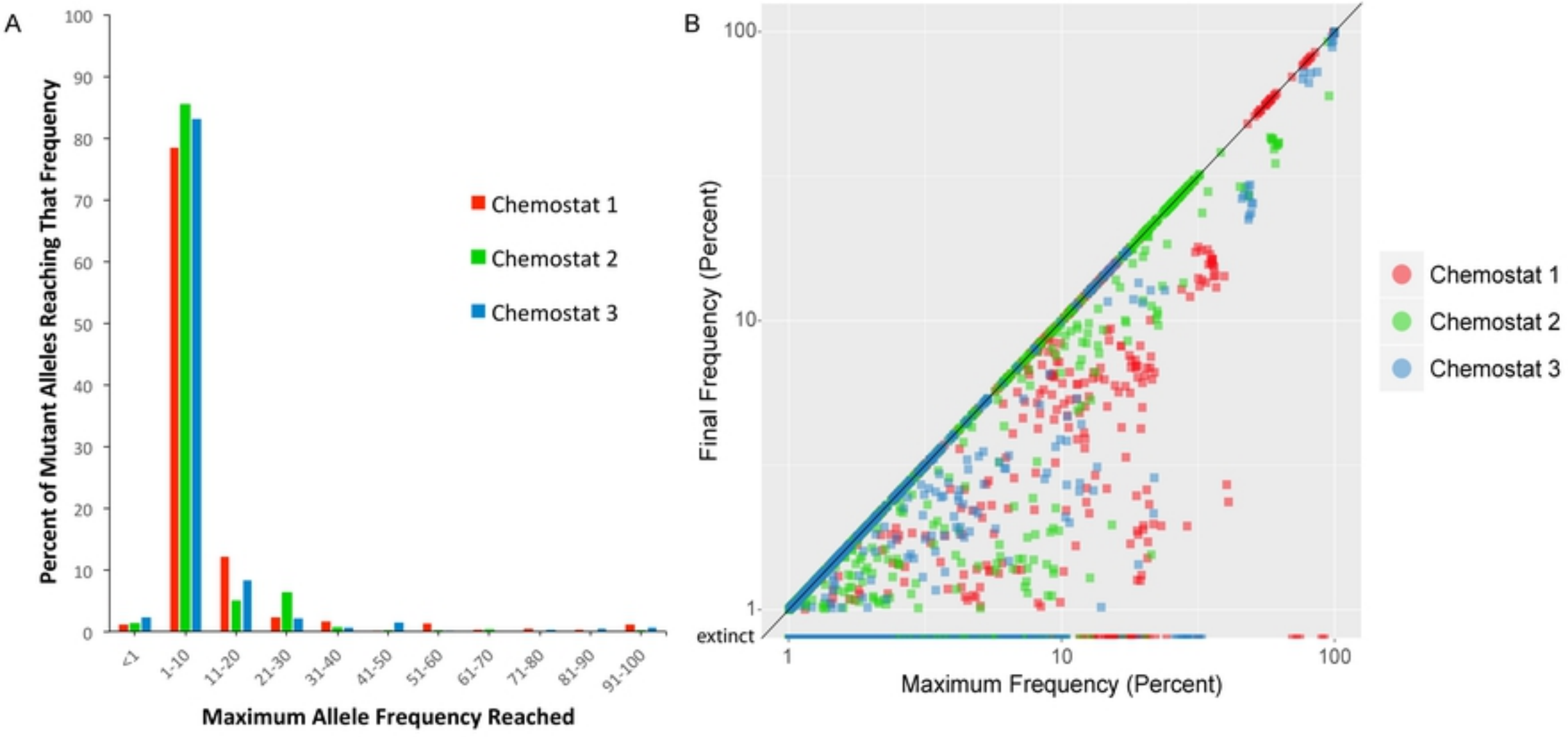

Fig. 2 Most de novo mutations only reach low allele frequencies

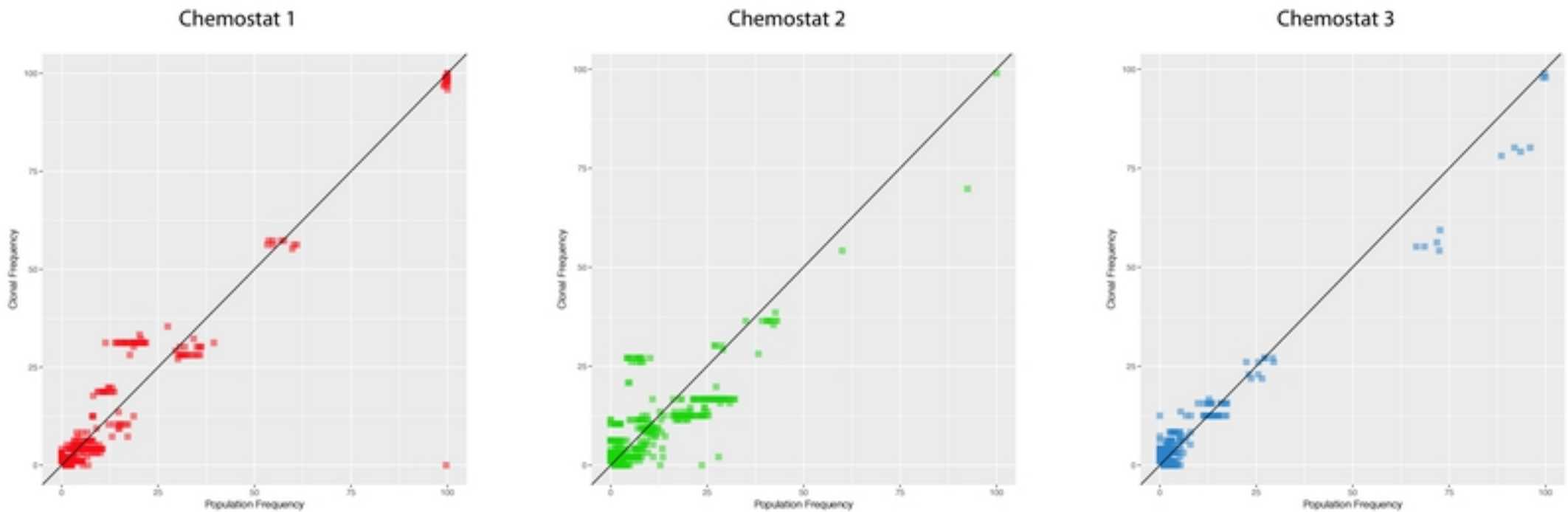

Fig. 3 Isolated clones are representative of the populations from

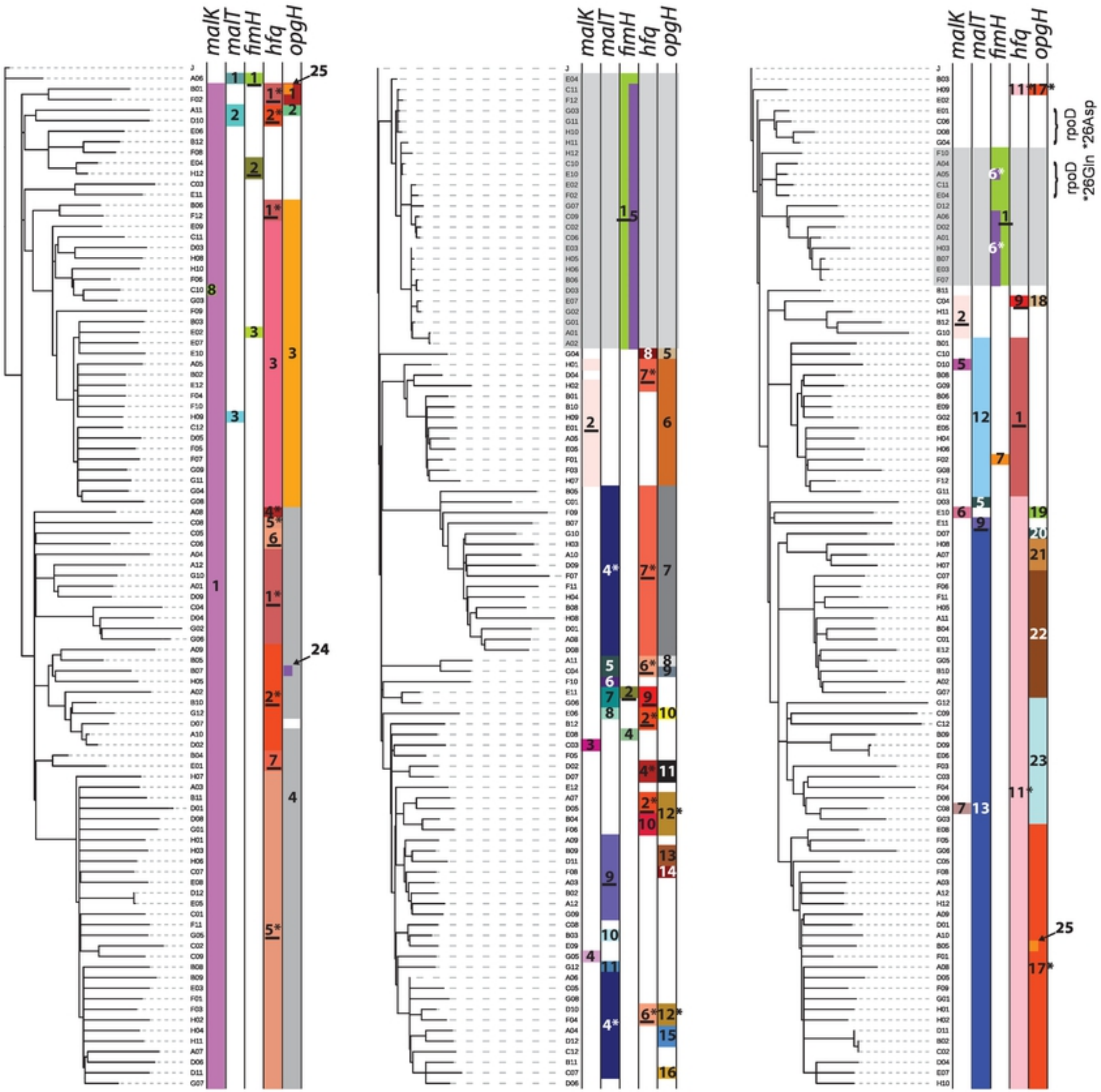

Fig. 4 Clone phylogenies

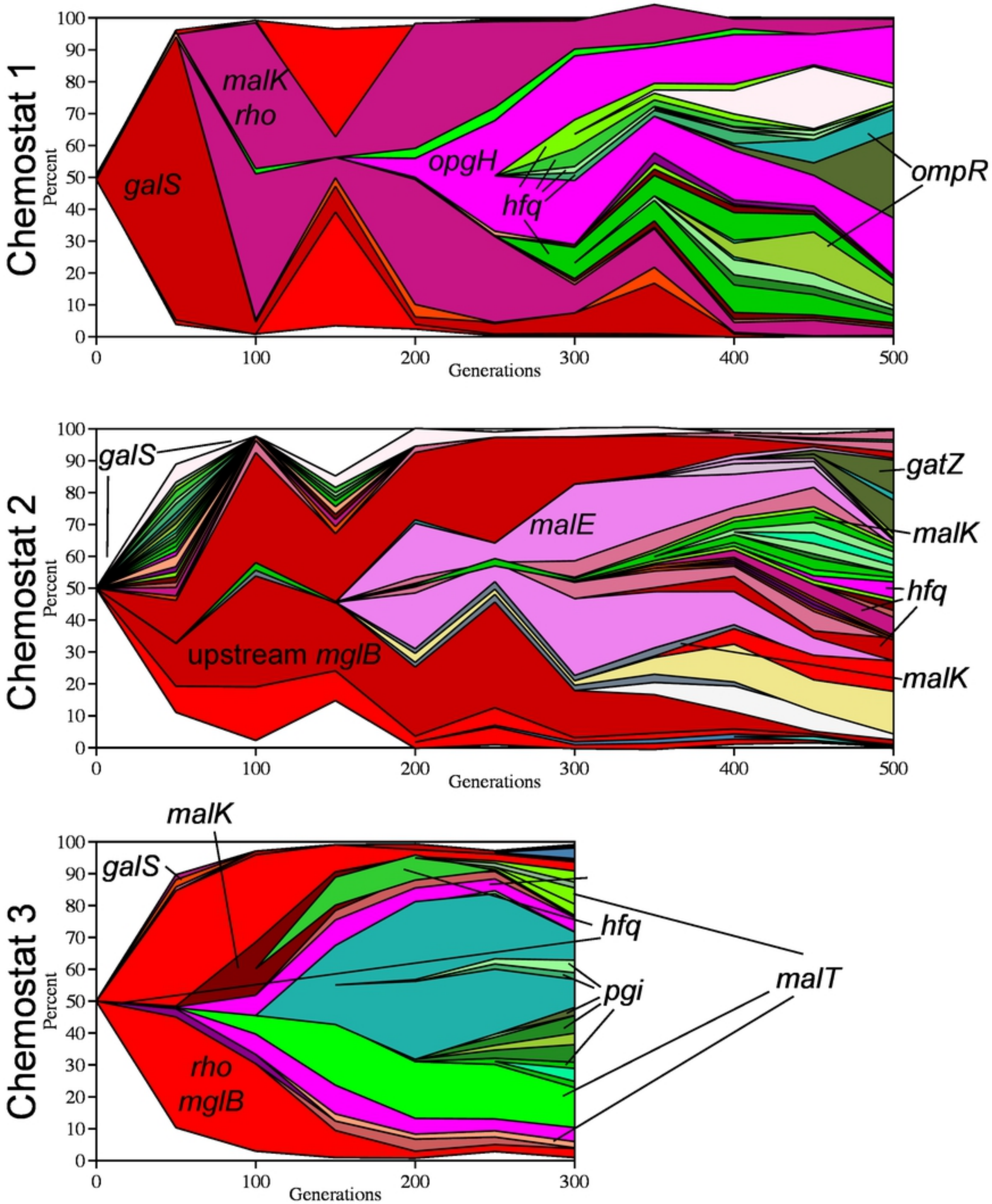

Fig. 5 Muller diagrams

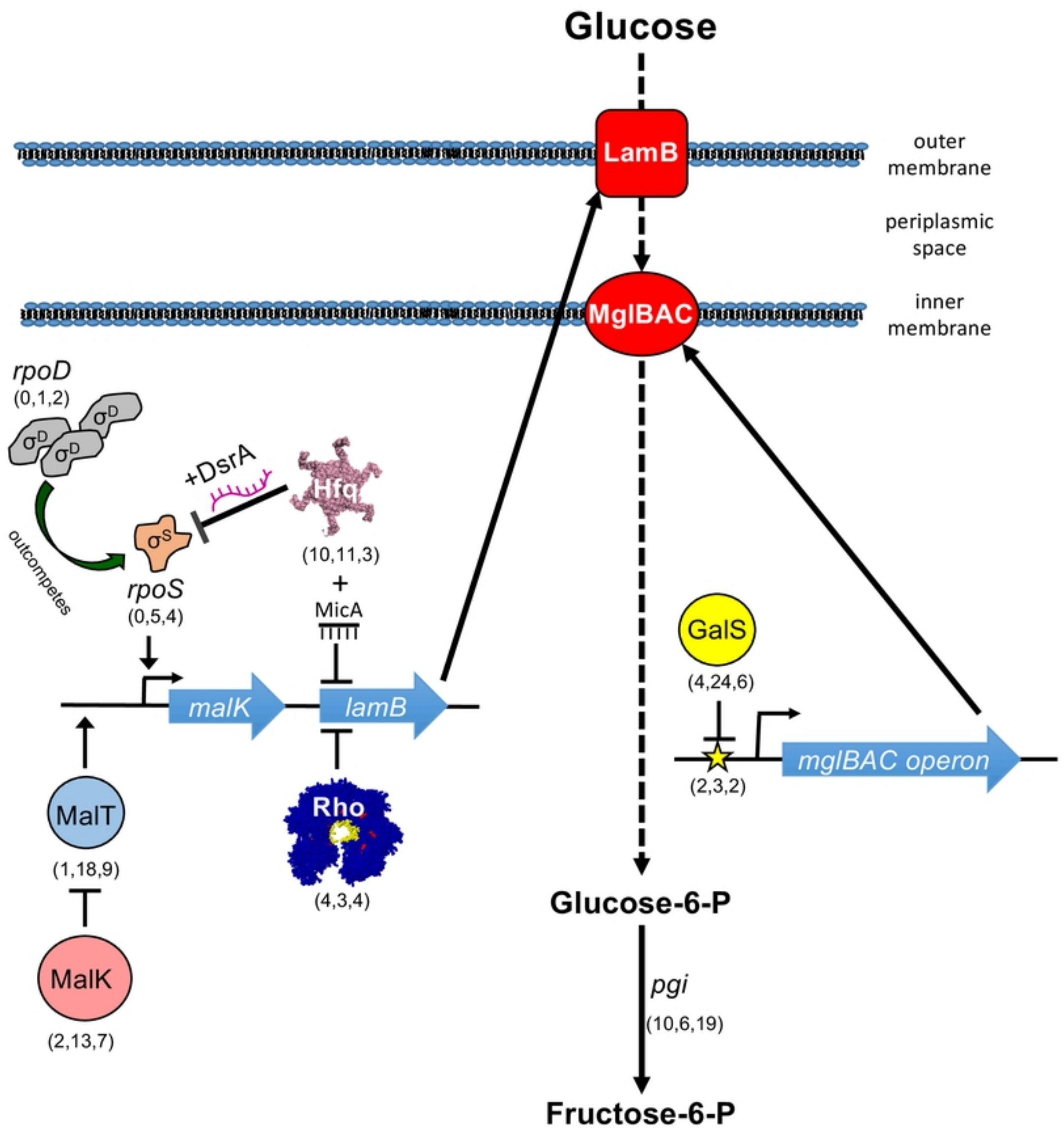

Fig. 6 Pathways containing frequently mutated genes