Supplementary Information

December 12, 2014

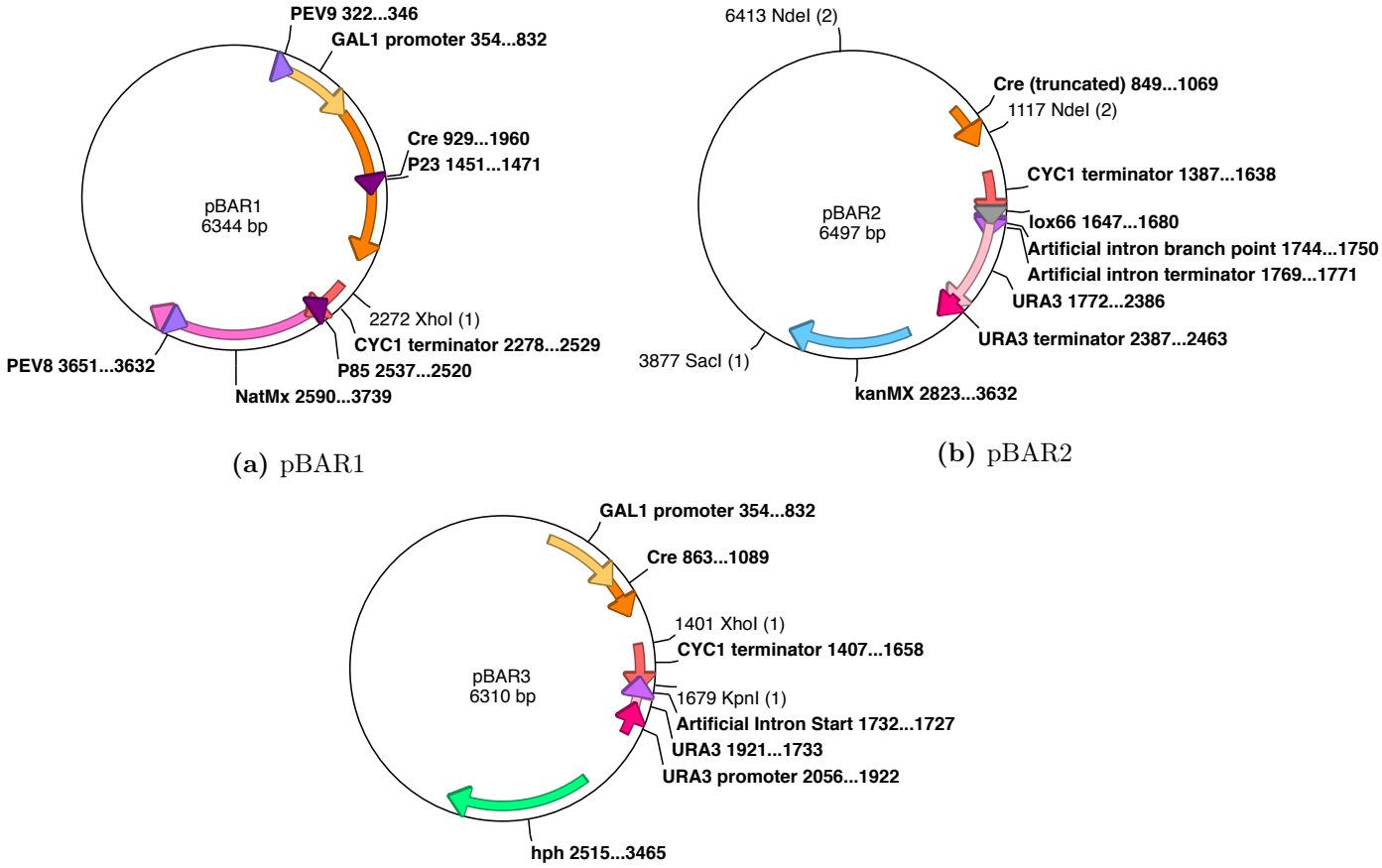
Contents

Experimental methods

1 Library construction, evolution and sequencing barcodes

1.1 Plasmid Cloning

Plasmids pBAR1 (Figure 1a), pBAR2 (Figure 1b) and pBAR3 (Figure 1c) were cloned from the following sources (all available from EUROSCARF) by standard methods: 1) plasmid backbone / bacterial origin from pAG32, 2) natMX, kanMX, and hygMX [1] from pAG25, pUG6, and pAG32 respectively, 3) Gal-Cre from pSH63 , 4) *URA3* from pSH47, 5) artificial intron, random barcodes and loxP sites were synthesized de novo (IDT).



(a) pBAR1

(b) pBAR2

(c) pBAR3

Figure 1: Maps of plasmids used in this study

1.2 Plasmid random barcode library construction

Random barcodes were inserted into pBAR3 by ligation. A primer containing a KpnI restriction site, a random 20 nucleotides, a lox71 [2,3] site and a region of homology to pBAR3 were ordered from IDT:

P85 = `CCAGCTGGTACCNNNNNAANNNNNTTNNNNNTTNNNNNATAACTTCGTATAGCATACATTATACGAACGGTAGGCGCGCCGGCCGCAAAT`

Random sequences were limited to 5 nucleotide stretches to prevent the inadvertent generation of restriction sites. The P85 and

P23 = `GCCGAAATTGCCAGGATCAGG`

primers were used to amplify a portion of pBAR1. Both the PCR product and pBAR3 were cut with KpnI and XhoI restriction sites and ligated together to generate plasmids containing a lox71 site and a random barcode. Ligation products were inserted into DH10B cells (Life Technologies) by electroporation, allowed

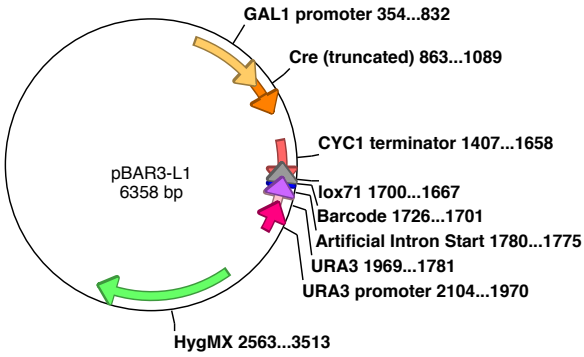


Figure 2: pBAR3-L1map

to recover from electroporation in liquid media for 30 minutes, and plated onto 180 LB-Ampicillin plates at a density of 3500 CFU/plate, a total of 630,000 colonies. During the recovery period in liquid media, some fraction of the cells could have undergone a cell cycle, meaning that our true library complexity is likely to be less than the number of colonies we observe. Colonies were pooled in 900 ml LB-Ampicillin and a fraction of the pool was used directly for plasmid preps to generate the plasmid library (pBAR3-L1) (Figure 2).

1.3 Yeast lineage tag library construction

The barcode "landing pad" was inserted by replacing the YBR209W dubious open reading frame in an S288C derivative, BY4709 with sequences derived from pBAR1 and pBAR2 by sequential homologous recombination. Disruption of YBR209W has been previously demonstrated to have no impact on fitness [4]. Sequential homologous recombination events were required because we found bacteria are unable to tolerate a plasmid that contains both Gal-Cre and loxP sites. First, pBAR1 sequence was amplified with the following primers:

PEV8 = GTTCTTTGCTTTTTTTCCCCAACGACGTCGAACACATTAGTCCTACGCACTTAACTTCGCATCTG

PEV9 = GCTTGCGCTAACTGCGAACAGAGTGCCCTATGAAATAGGGGAATGCATATCATACGTAATGCTCAACCTT

where underlined sequence correspond to sequences flanking the dubious open reading frame, YBR209W. The PCR product, containing Gal-Cre and the NatMX selectable marker, was inserted into the genome by homologous recombination [5] to create the SHA118 strain (MATα, ura3Δ0, ybr209w::Gal-Cre-NatMX). pBAR2 was cut with SacI and NdeI restriction enzymes and a gel fragment was isolated and transformed into SHA118 to create SHA185. The fragment contains homologous ends to replace the genomic NatMX marker with lox66 [2,3], half of an artificial intron and the 3' half of the *URA3* selectable marker [6].

SHA185 was transformed with the pBAR3-L1 barcode library and plated on galactose synthetic complete dropout plates lacking uracil. The galactose promotes Gal-Cre-induced recombination between the partially-crippled genomic lox66 and plasmid lox71 sites, causing insertion of the plasmid into the genome, and completing the *URA3* selectable marker. Insertion of the plasmid creates two genomic loxP sites: a fully functional loxP site located in the *URA3* artificial intron and a fully crippled lox66/71 [2,3] distant from the *URA3*. Because transformants contain only a single functional loxP site, insertion is is unlikely to easily reverse. Selection for *URA3*, and thereby the barcode residing in its artificial intron, is maintained during the evolutions by growing in media lacking uracil. Because the plasmid library contains a fixed number of barcodes (determined approximately by the number of bacterial insertion events), under-sampling barcodes could result in some barcodes missing or could create large differences in the frequencies of each barcode in the population, a problem we wished to avoid in our evolutions. We therefore plated $\sim 5 \times 10^6$ CFU,

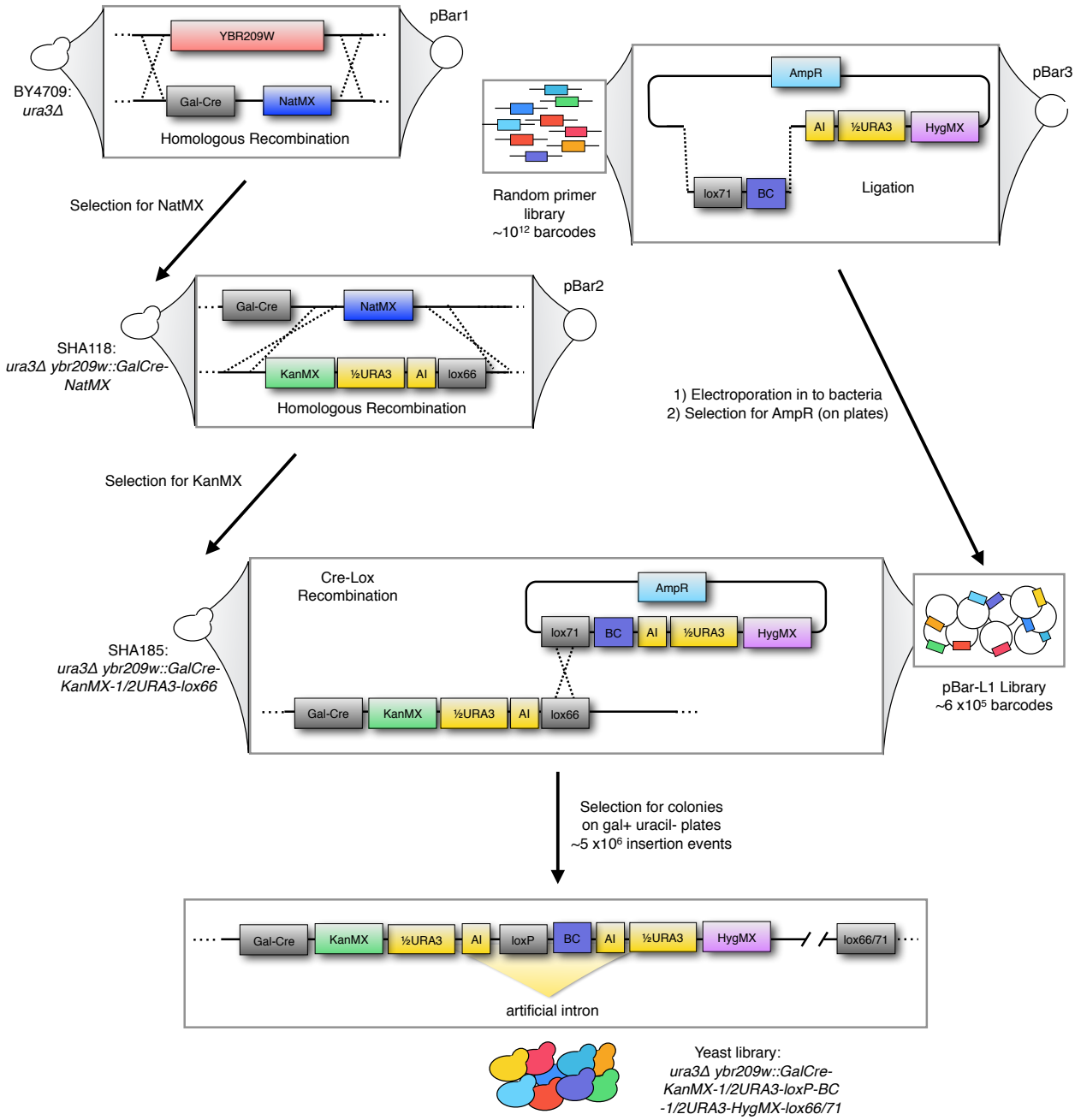


Figure 3: Schematic of barcode insertion

resulting in $\sim$ 10 insertion events per barcode for an estimated plasmid complexity of $\sim 5 \times 10^6$. LoxP insertion is highly efficient, allowing for this number of insertion events on $\sim$ 240 plates. Colonies from all plates were pooled and stored at -80° in 15% glycerol until beginning the evolutions.

1.4 Experimental Evolution

The lineage tag library was evolved by serial batch culture under carbon limitation in 100 ml of 5x Delft media [7] with 4% ammonium sulfate and 1.5% dextrose. Cells were grown in 500 ml Delong flasks (Bellco) at 30°C and 223 RPM for 48 hours between each bottleneck. Bottlenecks were performed by adding 400 μl of the evolution to fresh media. Cell counts were performed at each bottleneck to estimate the generation time. Contamination checks for bacteria or other non-yeast microbes were performed regularly. On the final time point of each evolution, we sexed several clones to assure that mating had not occurred during the evolution with an exogenous strain.

1.5 Sequencing Sample Preparation

Genomic DNA was prepared by spooling, as described [8]. A two-step directed PCR was used to amplify the lineage tags for sequencing. Because a small fraction of the total genomic DNA yields a PCR product ($\sim$100 base pairs out of a 12 MB genome size), we amplified 14.4 μg of template per sample, which corresponds to $\sim 10^9$ genomes or $\sim$ 2000 copies per unique lineage tag at time zero in the evolutions. First, a 3-cycle PCR with OneTaq polymerase (New England Biolabs) was performed in 24 reaction tubes, with $\sim$600 ng of template and 50 μL total volume per tube. Primers for this reaction were:

```
ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNXXXXXTTAATATGGACTAAAGGAGGCTTTT
```

and

```
CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNXXXXXXXXXTCGAATTCAAGCTTAGATCTGATA
```

The Ns in these sequences correspond to any random nucleotide and are used in the downstream analysis to remove skew in the counts caused by PCR jack-potting (see Lineage Tag Counts). The Xs correspond to a one of several multiplexing tags, which allows different samples to be distinguished when loaded on the same sequencing flow cell. PCR product was pooled into 4 pools of 50μL using 4 PCR Cleanup columns (Qiagen) at 6 PCR reactions per column. A second 24-cycle PCR was performed with high-fidelity PimestarMAX polymerase (Takara) in 12 reaction tubes, with 15 ul of cleaned product from the first PCR as template and 50 μL total volume per tube. Primers for this reaction were the standard Illumina paired-end ligation primers:

```
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
```

and

```
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT
```

PCR product from all reaction tubes was pooled into 50 μL using a PCR Cleanup column (Qiagen). The appropriate PCR band was isolated by E-Gel agarose gel electrophoresis (Life Technologies) and quantitated by Bioanalyzer (Agilent) and Qubit fluorometry (Life Technologies).

1.6 Lineage Tag Counts

Paired-end sequencing was performed on an Illumina HiSeq 2000. Each flow cell contained 2 to 4 multiplexed time points and 25% random genomic DNA. The genomic DNA was necessary to increase the read complexity for proper calibration of the instrument. Sequences were analyzed using custom written

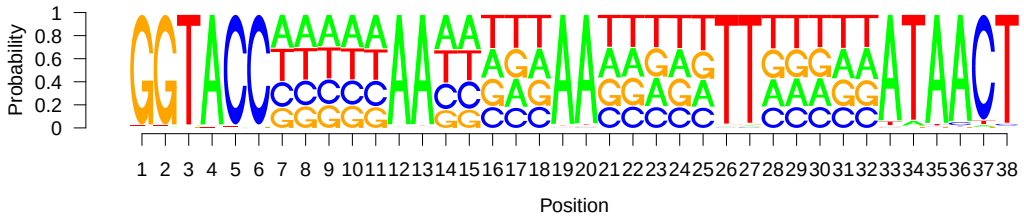


Figure 4: Position weight matrix for the barcodes inserted shows a relatively equal proportions of each nucleotide at each random position.

software in Python and R. Sequences were sorted by their multiplexing tags (the Xs in the primers above) and removed if they failed to pass two quality filters: 1) The average Illumina quality score for the lineage tag region must be greater than 30, and 2) the lineage tag region must match the regular expression `\D*?(.ACC|T.CC|TA.C|TAC.)\D{4,7}?AA\D{4,7}?AA\D{4,7}?TT\D{4,7}?(.TAA|A.AA|AT.A|ATA.)\D*`. We found a small fraction of barcodes with insertions or deletions in their randomized regions, which the regular expression encompasses. The regular expression also allows for one mismatch in the 4 bases on either side of the barcode region. One possible caveat is that a barcode that is excluded by our regular expression is present in the population and acquires a large adaptive mutation, causing clonal interference that is invisible to our sequencing assay. However, agreement between the change in mean fitness inferred from neutral and beneficial lineages (Figure 2, main text) suggests that we are accurately capturing the full extent of clonal interference.

Reads from the sequencing runs of all time points from the two evolutions were pooled and the number of occurrences of each unique read of the lineage tag region was counted (24,214,583 unique sequences of 1,576,711,485 total reads). We expected that the vast majority of unique sequences did not represent a true lineage tag, but rather sequences with a small number of mismatches from a true lineage tag, caused by PCR or sequencing errors.

We next clustered similar sequences to generate a set of lineage tag clusters, which are likely to consist of one true lineage tag (at a high frequency) and many similar sequence reads (at low frequency), generally with one or two mismatches. To generate clusters, we first considered only sequences with greater than 10 reads. We pairwise blasted (word size = 11, reward = 1, penalty = -2) each of these sequences against every other sequence. A sequence and all sequences that blast at an $e < 10^{-10}$ ($\sim$ 2 mismatches) to that sequence formed a cluster seed. Two clusters were joined if any any member (sequence) was present in both clusters. Cluster joining was repeated until the cluster number was stable. Sequences with less than 10 reads total were then matched to existing clusters by blast, using the same criteria. This method yielded 487,922 lineage tag clusters. The arbitrary, but computationally necessary, cutoff of a minimum of 10 reads to seed a cluster is likely to miss a few true lineage tags that begin at low frequency in the population and never rise over the course of either evolution (either by drift or accumulation of a beneficial mutation). However, reads that pass our filters (above) and do not eventually map to any lineage tag cluster constitute a small fraction of the total reads (0.17%), do not vary greatly across time points in our evolutions (0.15% – 0.19% of reads for individual time points), do not increase systematically in later time points, and are therefore unlikely to change the conclusions of this study.

Our estimated number of 487,922 true lineage tags is roughly consistent with the the number expected based on the complexity of our plasmid library. We counted 630,000 bacterial colonies during the plasmid ligation (see Plasmid Random Barcode Library Construction). However, because colonies were allowed to recover, and in some cases double, in liquid media for 30 minutes before prior to plating, a portion of bacterial colonies are likely to contain the same barcode. Our results suggest that 27% of bacteria doubled between electroporation and plating.

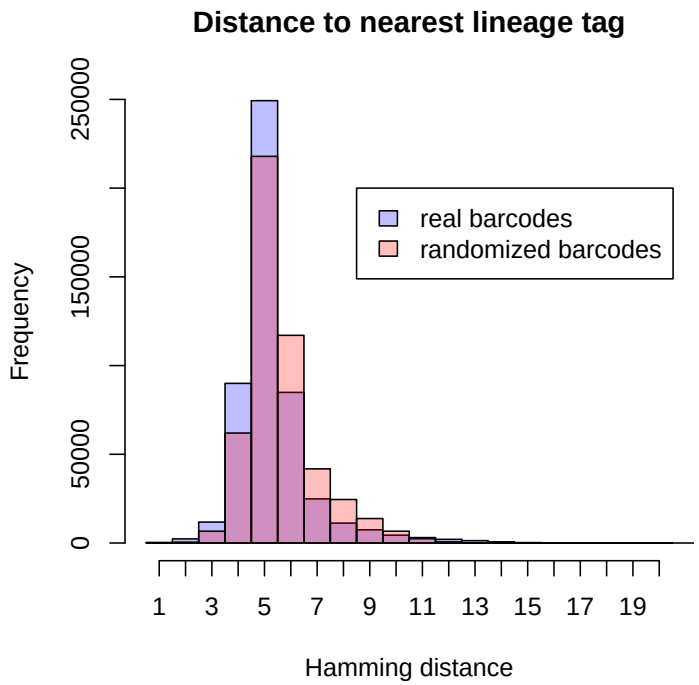


Figure 5: The distribution of hamming distances between the barcodes inserted (blue) and the expected distribution were they randomly placed in hamming space (red).

For each lineage tag cluster, we generated a consensus sequence by taking the most frequent base at each position across all members of the cluster. We refer these consensus sequences as "lineage tags" in this study. A position weight matrix of all lineage tags (Figure 4) shows that most randomized positions have a relatively even proportion of each base. Constant regions at positions 1 to 6, 12, 13, 19, 20, 26, 27, and 33 to 38 are dominated by the expected base. However, some variability does exist in the constant regions, due most likely to substitution or frameshift errors during the synthesis of the primers or PCR. To determine if our library of lineage tags conform to theoretical expectations given the frequency of bases at each position, we calculated the Hamming distance of each lineage tag to its nearest neighbor and compared this distribution to that when the base at each position is randomized across all reads (Figure 5). Distributions are similar, although randomizing bases results in slightly greater nearest neighbor distances, suggesting that a large fraction of the variability in the constant regions is due to frameshifts. That is, preserving the frame results in closer nearest neighbor distances than disrupting the frame through randomization.

One possible caveat of using random barcodes is that sequencing errors from an abundant lineage tag could erroneously contribute to read counts of a lineage tag with a similar sequence. This could result in "piggy-backing": as an adapted lineage rises in frequency, closely-spaced lineage tags also rise in frequency and are erroneously interpreted as containing a beneficial mutation. Yet, we find that the vast majority of lineages tags are at least 4 mismatches away from a nearest neighbor (97%, Figure 5). Thus, a sequencing read is unlikely to be assigned to the the wrong lineage tag unless it contains 3 or more errors. By measuring the frequency of reads that do not match exactly to any lineage tag, we conservatively estimate the per-base error rate of our experiment to be 0.18%, consistent with previous reports [9,10]. Thus, a conservative estimate of our rate of single base error per read is 6%. At this rate, the frequency of reads with 3 or more errors in the 20 random nucleotides of each barcode ($\sim 2 \times 10^{-4}$) does not constitute a large enough fraction of the reads to impact our results. For example, an adaptive lineage that expands to 6×10^6 cells ($\sim$1% of the population in our experiments, the largest lineages we observe) would yield only $\sim$ 1300 triple

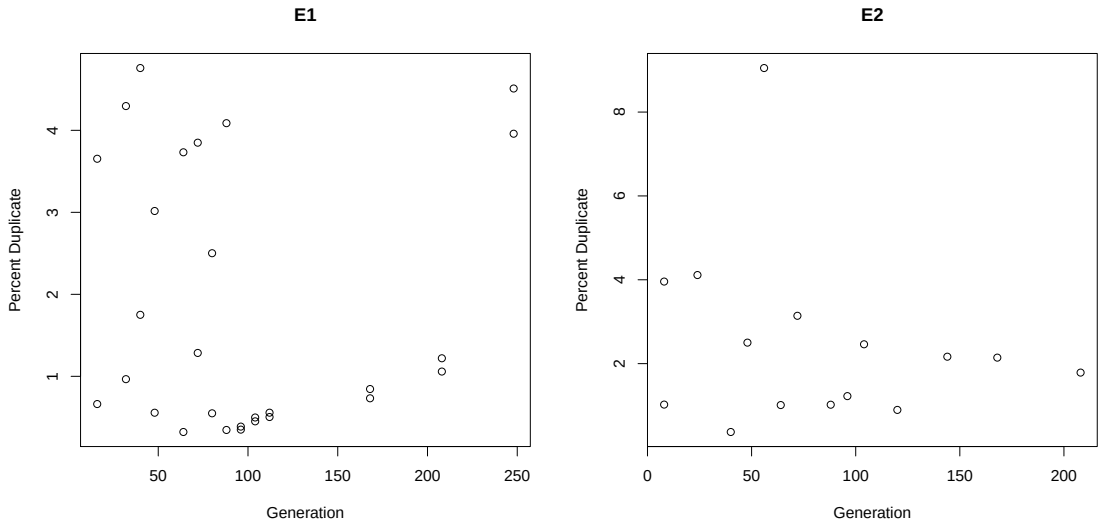


Figure 6: Duplicates per generation for each replicate

errors, with only a handful of these reads erroneously mapping to to the same incorrect lineage tag.

For each time point, we matched all sequencing reads onto our set of lineage tag clusters as an initial count of the number of occurrences of each lineage tag. One source of bias in these counts that we wanted to avoid is PCR jack-potting or other non-linearities between the amount of template for a lineage tag and the number of sequences observed for that tag. We removed these errors by attaching two random 8mers to each template molecule in the first few rounds of PCR (see Sequencing Sample Preparation). Because the total sequence space of two random 8mers is large ($\sim 4 \times 10^9$ possibilities), it is unlikely that any two template molecules from the same time point that contain the same lineage tag will be attached to identical pairs of 8mers. Thus, sequence reads for a lineage tag that contained the same pair of 8mers were counted as PCR duplicates and removed from our final counts. Overall, the extent of PCR duplicates was nominal in our experiments (less than 2.4% of all reads that match a lineage tag cluster), suggesting that template abundance and primer annealing efficiency was sufficient in our protocol. We found no evidence that the number of PCR duplicates changed systematically over the course of either experiment (Figure 6) or that lineage tags that became abundant contained more duplicates as might be expected if the number of random 8mer template tags were limiting.

1.7 Fluctuation Test

Fluctuation tests for resistance to canavanine due to mutations in the CAN1 locus were performed as described in [11], except all growth was performed in 5x Delft media with 4% ammonium sulfate and 1.5% dextrose. The per base per generation mutation rate of the barcode library (3.66×10^{-10}) is similar to an S288C control strain (4×10^{-10}, this study), and previous reported estimates for S. cerevisiae (1.73×10^{-10} to 6.44×10^{-10}) [11, 12]. Fluctuation tests from three different colonies (and presumably different barcodes) of the barcode library have similar per base per generation mutation rates ($4.4 \times 10^{-10}, 4.21 \times 10^{-10}, 2.37 \times 10^{-10}$).

2 Validation of fitness measurements using fluorescent labels

The fitness of clones that were identified as adaptive in replicate E2 were independently validated as follows. BY4709, which is the ancestral strain to the barcode library, was transformed with the plasmid pGS63 [4]. This generates an ancestral clone with a genomically-integrated YFP construct to use as a fluorescent

reporter when conducting fitness assays against the non-fluorescent adaptive clones. Multiple YFP tagged transformants were competed against neutral lineages to ensure they had ancestral fitness and had not picked up a deleterious mutation during the transformation process before using them in competition against adaptive clones. Additionally, clones were isolated from the frozen stock of generation 88 of the replicate E2. The barcode on each clone was Sanger sequenced for identification, and the expected fitness of each clone was identified from the population sequencing analysis. We chose 26 putatively adaptive clones and 5 putatively neutral clones from this sample for fitness validation.

Each of the test clones and the YFP clone were streaked out from freezer stock onto 5x Delft [7] agar plates and grown for 3 days. Colonies were then inoculated into 3mL liquid 5x Delft media [7] with 4% ammonium sulfate and 1.5% dextrose and grown in test tubes in a roller drum at 30C for 2 days, after which 400ul was transferred to 100mL of the same media in 500mL Delong flasks (Bellco) and grown in a platform shaker at 30C for 2 days at 223 rpm. This final set of conditions is identical to the conditions the strain was evolved in. 80μl of each test clone was mixed with 720μl of the YFP clone in eppindorf tubes. One additional tube containing no test clones and 800μl YFP culture was also made. 400μl of each mixture (including the YFP only culture) was inoculated into evolution condition flasks, and the remaining 400μl was used for flow cytometry analysis. Each flask was grown for 2 days, after which 400μl of each culture was transferred to fresh flasks and a further 400μl was used for cytometry analysis. This protocol was repeated for a total of 3 transfers. A final cytometry time point was taken 2 days after the last transfer, resulting in five total time points (the initial mixture, at each of the three transfers, and after the last transfer reaches saturation) covering 32 generations (8 generations per 2 day growth phase).

For each cytometry time point, the 400μl samples were inoculated into 2ml of 5x Delft media [7] with 4% ammonium sulfate and 1.5% dextrose and grown in test tubes on the roller drum for 2 hours. 50μl of each test tube culture was diluted into 450μl of sterile water, which was then used for cytometry analysis. Cytometry-ready samples were kept at 4C for up to 4 hours while waiting for machine access. Each test sample was analyzed for 10,000 events on a BD-FACScan analyzer (Stanford, Stanford Shared FACS Facility) which was calibrated with beads before each use. Events which reached the maximum FSC value were not counted, as they likely represented multi-cell agglomerates. The pure YFP sample was similarly analyzed for 50,000 events. We utilized the BluFL1 (560nm short pass splitter, 525/50 band pass filter) and BluFL2 (640nm long pass splitter, 615/25 band pass filter) filters for the fluorescent analysis. After sampling, events that had FSC values less than 11 were excluded, as these likely represent non-living cells. For the events that passed this minimum criteria, we computed the percentage of fluorescent cells. Previous analysis with known ratios of fluorescent to nonfluorescent cells had established a threshold of

$$blu2 < blu1 \times 1.65 - 5 \qquad \text{and} \qquad blu1 > 6.5 \tag{1}$$

for differentiating fluorescent cells from nonfluorescent cells. As not all cells containing the fluorescent tag actually fluoresce, we use the pure YFP culture that was grown and sampled alongside the test cultures to calibrate the percentage of nonfluorescent YFP cells using a simple linear model independently for each sample time point.

The number of observed non-fluorescent cells that we expect to contain YFP = The observed number of fluorescent cells in the mixture $\div$ the percentage of pure YPF cells that fluoresce $\times$ the percentage of pure YFP cells that do not fluoresce. We then subtract the expected number of YFP non-fluorescent cells from the observed number of nonfluorescent cells to get the true number of nonfluorescent cells in the culture.

For each strain, the natural log of the percent of nonfluorescent cells was plotted against the number of generations that had elapsed by that time point (0, 8, 16, 24 and 32). The slope of the linear fit is the relative fitness advantage of the test strain compared to the YPF-tagged ancestral strain.

Theory and Data Analysis

3 Useful numbers and notation

What	Value	Symbol
Fitness effect of mutation	$\sim 2.5\% - 15\%$	s
Establishment time of mutation	$[-100, 50]$	τ_{est}
Occurrence time of mutation	$[-100, 50]$	τ_{mut}
Variance in offspring number(per generation)	≈ 3.5	$2c$
Cells at bottleneck	$\approx 7 \times 10^7$	N_b
Cells at saturation	$\approx 1.7 \times 10^{10}$	N_s
Effective population size	$\approx 6 \times 10^8$	N_e
Unique barcode sequences	$\approx 500,000$	L
Sequencing reads per time point	$\sim 3 \times 10^7$	R
Median initial lineage size (at bottleneck)	~ 100	n_b
Median initial lineage size (effective)	$\sim 10^3$	n_e
Range of lineage sizes (at bottleneck)	$\sim 10 - 600$	—
Mean reads per barcode	~ 50	$\bar{r}$
Generations between time points	≈ 8	T_{cyc}
Dilution factor at bottleneck	$1/250$	Δ
Generations of evolution	~ 150	—
Adaptive lineages	$\approx 25,000$ (E1, $t = 112$)	—
Adaptive lineages common across replicates	$\approx 6,000$	—
Mean fitness increase	$\approx 6\%$ (E1, $t = 112$)	$\bar{x}(t)$

Table 1: Useful numbers and symbols for quantities commonly used in the analysis.

Time point	0	8	16	24	32	40	48	56	64	72	80	88	96	104	112
Depth in E1 ($\times 10^6$)	222	–	84	–	77	51	86	–	83	63	58	54	48	56	53
Depth in E2 ($\times 10^6$)	222	58	–	55	–	26	30	51	30	22	–	20	32	–	–

Table 2: Read depths across time points for the two replicate evolutions. Only these time points are used in the analysis.

4 Statistical dynamics of beneficial mutations

4.1 Establishment probability and establishment time

A single beneficial mutation that has a (small) fitness advantage s relative to a constant mean fitness has on average $1+s$ offspring per generation. In the absence of variation in offspring number, the descendants of a single founder would grow exponentially in time as $(1+s)^t \approx e^{st}$. However, if there is some variation in offspring number around the mean, the dynamics is stochastic and the mutation sometimes fluctuates to extinction (drifts out). We consider a general case where the variance in offspring number is $2c$, with c a constant. If it does not go extinct, the mutant population will eventually reach large enough numbers to grow exponentially and essentially deterministically: i.e. it "establishes". The relative likelihood of these outcomes and the number of cells, $n(t)$, that share the beneficial mutation given that it survives, can be computed by solving for, and inverting, the moment generating function for a birth-death process of a single founding beneficial mutation with fitness effect s entering at time $t=0$ (see Section 13.1 and [13]). The probability of surviving drift and "establishing" in the population is

$$Pr(\text{establishing}) \approx s/c \tag{2}$$

with $2c$ equal to the (effective) variance in offspring number for a single cell per generation; the probability of extinction is thus $1-s/c$. At long times ($t \gg 1/s$) the mutant population grows as

$$n(t) \approx \nu \frac{c(e^{st}-1)}{s} \qquad \text{with the probability density of } \nu \qquad \rho(\nu) \approx \left(1-\frac{s}{c}\right)\delta(\nu) + \frac{s}{c}e^{-\nu}, \tag{3}$$

the delta-function part representing extinction.

Typically $s/c \ll 1$ so the majority of mutations that enter the population drift to extinction before establishing. If the beneficial mutation survives to reach n copies in the population, and since all cells with the mutation are independent, the probability of extinction becomes $(1-s/c)^n \sim e^{-ns/c}$. Therefore, if the mutant population reaches substantially more than $n \gtrsim c/s$ copies it is unlikely to go extinct:

$$n_{est} \sim c/s \tag{4}$$

is the typical size one a mutation must reach in order to establish. If it does establish, the fluctuations in $n(t)$ at later times will be small: the value of the coefficient ν is thus primarily determined by fluctuations before the mutant establishes. It is convenient to define the "establishment time", τ_{est}, as the time at which the adaptive mutant population would have had a size of $\sim c/s$ cells had it grown exponentially with no fluctuations from then on as shown in Fig 7A. From (3) this gives

$$\nu = e^{-s\tau} \qquad \text{and the probability density of } \tau \qquad \rho(\tau)d\tau = \exp\left(-s\tau - e^{-s\tau}\right) s\, d\tau. \tag{5}$$

For a mutation that enters the population with $\tau_{mut}=0$ the most likely establishment time is $\tau_{est}=0$, and the width of its distribution is $\sim 1/s$. The establishment time therefore roughly corresponds to the true time of occurrence of the mutation, τ_{mut}, with errors of $\pm 1/s$ generations as shown in Figure 7B. We point out however that the distribution is asymmetric: mutations with establishment times substantially earlier than $-1/s$ are far less likely than mutations with establishment times substantially later than $+1/s$.

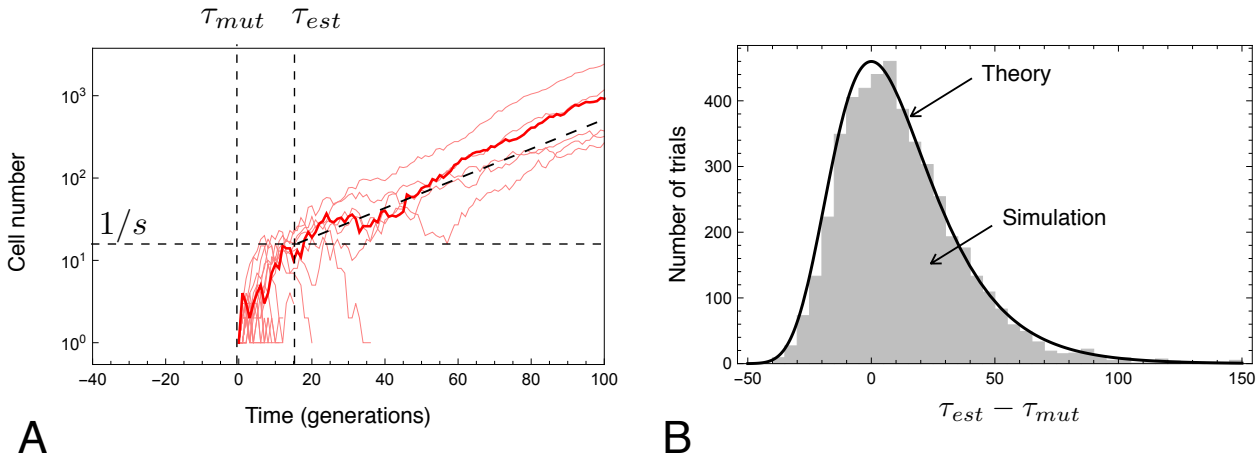


Figure 7: (A) 50 simulated trajectories (light red) of beneficial mutations with a fitness effect $s = 0.05$ that were present in a single copy at $t = 0$. Each generation a cell has a number of offspring drawn from a Poisson distribution with mean $(1 + s)$. The long term growth of cell number is of the form $n \approx (c/s)\exp(s(t - \tau_{est}))$, where $c = 0.5$ is half the variance in offspring number and s the fitness advantage. The typical trajectory (dark red) of a mutation destined to survive grows faster than exponentially during drift phase ($t \lesssim 1/s$). The establishment time, τ_{est} is the time at which the cells would have reached $1/s$ cells were they to have grown exponentially at rate s throughout their trajectory (angled dashed line intercepting horizontal dashed line). (B) A histogram (gray) of the establishment time τ_{est} for 50,000 simulated trajectories. The black curve is the theoretical expression from Eq. 5. The most likely τ_{est} is the occurrence time of the mutation τ_{mut} (zero in this example) with errors of $\pm 1/s$ that arise from the stochastic jumps taken while at low numbers. The distribution is asymmetric: establishment times earlier than $-1/s$ are less likely than establishment times later than $1/s$. The 95% confidence interval occurs between $-20 < \tau_{est} - \tau_{mut} < 50$ in this case.

4.2 Establishment time, τ_{est}, for mutations fed from a constant population

Consider now a constant population of n_0 neutral cells that can feed beneficial mutations — all with fitness advantage s — at a rate U_b. This is the process by which mutations enter our barcoded lineages. The total number of beneficial cells that are descendants of this initial subpopulation will grow exponentially as

$$n(t) \approx \frac{ce^{s(t-\tau_{est})}}{s} \tag{6}$$

however now the establishment time τ_{est} must contain information about how rapidly mutations — possibly many of them — enter as well as how they fluctuate once they enter. The statistics of τ_{est} in this case are obtained by solving the mutation-birth-death process as outlined in (13.1). The result is similar to that of a single mutation, with the distribution of τ_{est} altered to reflect the process of mutation

$$\rho(\tau)d\tau = \frac{sd\tau}{\Gamma(n_0U_b)} \exp\left(-n_0U_b(s/c)\tau - e^{-s\tau}\right). \tag{7}$$

The trajectories that rest from such a feeding process are shown in Figure 8A. Here a feeding population of $n_0 = 1000$ cells feeds mutations at a rate $U_b = 10^{-4}$ per cell per generation into a fitter class ($s = 0.05$) that, if successful in establishing, grow exponentially. Such a feeding process produces a broad distribution of establishment times Figure 8B. The median time can be calculated by asking when the cumulative number of reproductions (n_0t) multiplied by the probability of mutating and establishing ($U_b(s/c)$) is order 1/2. There are two important distinctions from the single mutant case in Figure 5. First, all trials will, given long enough time, succeed in obtaining an established mutation because the feeding population is constant and continues to feed regardless of the fate of the mutations (this explains why more lines in the figure increase exponentially). Second, the distribution of τ_{est} is highly asymmetric: while establishment times

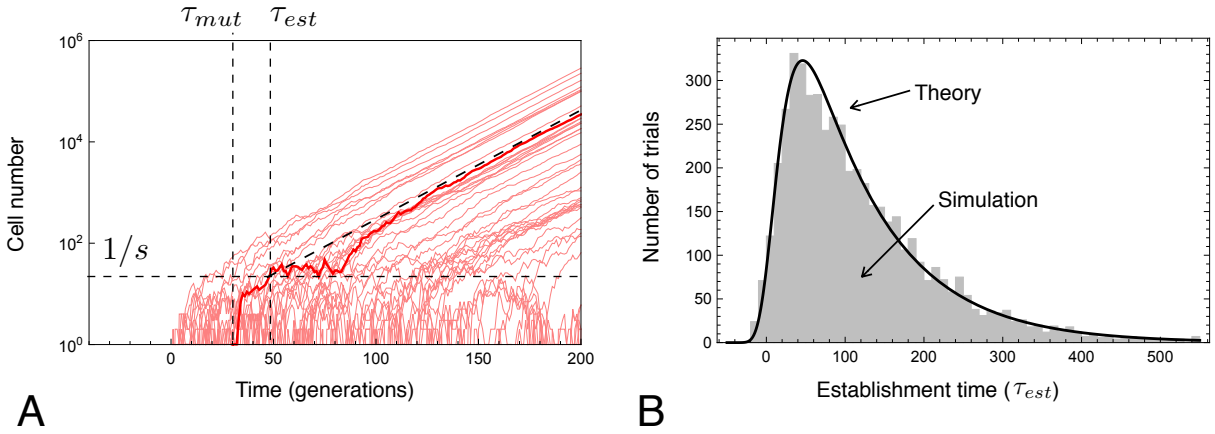


Figure 8: (A) 50 simulated trajectories of the number of cells with a beneficial mutation ($s = 0.05$) that occur via mutation from a pool of $n_0 = 1000$ cells at rate $U_b = 10^{-4}$. Each generation a cell has a number of offspring drawn from a Poisson distribution with mean $(1 + s)$. The long term growth of cell number is again of the form $n \approx (c/s)\exp(s(t - \tau_{est}))$, where $c = 0.5$ is half the variance in offspring number, s the fitness advantage. The establishment time, τ_{est} is again the time at which the cells would have reached $1/s$ cells were they to have grown exponentially at rate s throughout their trajectory (angled dashed line intercepting horizontal dashed line). (B) The distribution of establishment times from 5,000 simulations with the same parameters as (A). The black curve is the predicted distribution from Eqn. 7. The median time can be easily estimated by realizing the cumulative number of reproductions multiplied by the probability of establishment must be $\sim 1/2$ i.e. $n_0 U_b (s/c)t \approx 1/2$ which here gives $t \approx 50$ generations. At positive τ_{est} the distribution decays off with decay time $1/(n_0 U_b (s/c))$ which in this case is ≈ 100 generations.

earlier than $-1/s$ are unlikely, at positive establishment times the distribution decays exponentially with decay time $1/(n_0 U_b (s/c))$ which can be very long. In the above simulations the feeding rate was low enough that $n_0 U \ll 1$, and as one can see, the emergence of the mutants is stochastic. If however one considers the total population $NU \gg 1$ and the emergence of the first mutants becomes essentially deterministic as we now discuss.

4.3 The distribution of mutation rates $\mu(s)$, the deterministic approximation, and, "predominant" s

The previous discussion assumed all mutations were of the same effect size s. In reality there will be different mutation rates to different fitness effects s, which we capture with the distribution $\mu(s)$:

$$\text{Mutation rate to range } [s, s + ds] = \mu(s)ds \tag{8}$$

the total beneficial mutation rate is then

$$\int_0^\infty \mu(s)ds = U_b \tag{9}$$

however, as we discuss in the following paragraph, the total beneficial mutation rate can be a misleading quantity. The reason it can be misleading is that not all beneficial mutations are "created equal". The range of fitness effects that drive the increase in mean fitness (called "predominant" s range) can be a small fraction of the total — especially at later times. To see why this is the case we consider the growth of cells in fitness range $[s, s + ds]$.

Deterministic approximation. Consider the constant feeding process of the previous section again. The growth of the mutant class is determined by two things: the fitness of the mutant, s, and the number of independent mutations that contribute to the expansion $\approx N\mu(s)ds$ (rationale: there is a window of $\sim (c/s)$ generations in which mutations contribute significantly to the expansion, mutations establish at a

rate $N\mu(s)ds(s/c)$ during this time so that the product is $N\mu(s)ds$). If the population size is large enough then many independent mutations contribute to the expansion

$$N\mu(s)ds \gg 1 \tag{10}$$

then the stochasticity is small and the expansion of the class is close to deterministic (the reason being one is averaging over many mutations, that are individually stochastic, but collectively almost deterministic). In the case where the expansion of cells with fitness in the range $[s, s+ds]$ is deterministic, the fraction of the total population $f(ds,t)$ with fitness in the range $[s, s+ds]$ is easily calculated

$$f(ds,t) = \frac{\mu(s)ds}{s}[e^{st} - 1]. \tag{11}$$

This is an important result, and is used later to infer the spectrum of mutation rate as a function of fitness. It links a quantity of interest (the distribution of mutation rates across fitness effects) to an easily measured quantity (the fraction of cells in a given fitness range). Notice that it does not depend on the details of the growth process (e.g. population size, N, or variance in offspring number, $2c$) because these quantities do not affect the mean growth.

A key insight from this deterministic approximation is the importance of the product

$$\mu(s)e^{st} \tag{12}$$

in determining which fitness effect is most abundant in the population, and therefore which range of s actually matters in driving the increase in fitness. Mutations with small fitness effects might be more common, but because they grow exponentially more slowly, they are rapidly outcompeted by rarer large effect mutations. How long this takes depends on the shape of $\mu(s)$, however one can use this argument to define the "predominant" fitness effect $\tilde{s}(t)$ that is most abundant in the population at time t and the mutation rate to this $\tilde{s}$ which we denote as $\tilde{U}(t)$, both of which are functions of time (see Section 11.1 for details). A simple example is if the distribution of mutation rates to fitness effects falls off as a Gaussian, $\mu(s) \sim \exp(-(1/2)(s/\lambda)^2)$. Then exponent in the product in 12 becomes $(1/2)(s/\lambda)^2 + st$ which is maximized at

$$\tilde{s} = \lambda^2 t \tag{13}$$

so that the predominant fitness effect in this case increases linearly in time. The deterministic approximation only holds when many mutations contribute to the expansion of a fitness class, i.e. $N\mu(s)ds \gg 1$. At late times, the value of $\tilde{s}(t)$ will be large enough that the mutation rate to it ($\tilde{U}(t)$) will be small enough that $N\tilde{U}(t) \lesssim 1$. Beyond this time the dynamics are stochastic, because only of order one mutation drives the mean fitness in each fitness class. In the experiment we observe this transition: from a deterministic expansion of a large number of mutations early on to a more stochastic expansion of a few mutations later. We discuss this effect in more detail in Section 11.1.

4.4 Offspring distribution through a cycle

In the previous section we saw that, provided $N\mu(s)ds \gg 1$, the total number of mutant *cells* in a given fitness range is independent of the population size, changes in population size over time, or details of the growth-process of cells e.g. variance in offspring number. The same is *not true* if one is interested in the number of *mutations* that establish or on size of neutral fluctuations. The probability of mutations establishing in the population and the size of neutral fluctuations both depend on the details of the population size over time and on the variance in offspring number.

Variance in offspring number and neutral fluctuations. In our experiment cells are bottlenecked and then grown up through $T \sim 8$ generations. The average number of offspring for a single neutral cell through this cycle is one. The variance in offspring number through the cycle we denote by $2c \approx 3.5$ (determined

in 5.4). This variance has contributions from both Poisson noise of the bottleneck (≈ 1) and from the variations in growth through the cycle (≈ 2.5). As we show below this variance is larger than would be expected by purely Poisson noise sampling each generation and is likely to have a substantial contribution from variations in the lag time to start dividing after dilution into fresh medium: these are typically of order one generation.

Although we can only measure the variance in offspring number through a cycle, one can ask what our measured values for $2c = 3.5$ corresponds to for the average variance in offspring number per generation. Consider each doubling, where the expected number of cells increases by a factor of two. If there were Poisson noise each generation, the variance in number of offspring in the first generation is 2 but this grows up through $T-1$ generations. The variance in the second generation is 4, which is scaled up though $T-2$ generations and so on. Including the bottleneck of $\Delta = 2^{-T}$ at the end of the cycle we see that the variance in cell number across the entire cycle is

$$var(\text{cycle of } T \text{ generations }) \approx \Delta^2 \left(2 \times 2^{2(T-1)} + 4 \times 2^{2(T-2)} + 8 \times 2^{2(T-3)} + ...256 \times 2^{2(T-8)}\right) + 1 \quad (14)$$

$$\approx (1/2 + 1/4 + 1/8 + ...1/256) + 1 \quad (15)$$

$$\approx 2 \quad (16)$$

where the last term in the first two lines is the noise introduced at the bottleneck. The variance in offspring number of a single cell per generation assuming Poisson noise each generation is therefore of order $2/T$, whereas our measured value would be $3.5/T$. The scaling with $1/T$ is important. Because cells are permitted to expand exponentially for T generations, neutral fluctuations *per generation* are smaller by a factor of $1/T$, which in our case is significant. One can think of this either as a rescaling of the effective variance in offspring number

$$c \to c/T \quad (17)$$

or equivalently as a rescaling of the population size

$$N_b \to N_e = N_b T, \quad (18)$$

the conventional *effective population size* for drift.

Establishment of de novo beneficial mutations. The rate at which mutations establish in the population also depends on T. The probability of a beneficial mutation entering during the first doubling of the cycle and surviving the bottleneck is $\approx N_b U_b$. It is twice as likely to occur during the second doubling, but half as likely to survive the bottleneck, so the probability of a mutation arising at some point in the cycle and being present as a single cell at the start of the next cycle is largely independent of where in the cycle it arises, always being proportional to $N_b U_b$. Consider now a beneficial mutation present in a single copy at the beginning of a cycle. Through the cycle the mean increase in abundance is $\exp(sT) \approx 1 + sT$ with the variance (see above) of order c. Using the expression from Eq. 2 we therefore see that, provided $sT \ll 1$ the establishment probability of a mutation becomes

$$Pr(\text{establishment}) \approx \frac{2 \times \text{Mean growth rate}}{\text{variance}} \approx \frac{sT}{c} \quad (19)$$

which is consistent with our previous rescaling of $c \to c/T$. It should be noted however that this approximation breaks down if $sT \sim 1$. The establishment probability for mutations with $s > 1/T$ that are present at the beginning of the growth cycle approaches unity. Under our growth cycle, because variance is reduced by a factor of T, de novo mutations establish with a probability that is scaled up by a factor of T. However, for the same reason, the size of any given mutation is typically reduced by a factor of T. The expected number of *cells*, which is the product of the number of mutations and their population size is therefore independent of the rescaling, which explains why in the deterministic approximation these details do not enter. To verify the above we performed simulations with different cycle lengths T and measured the number of mutations entering the population, the results are shown in Figure 9.

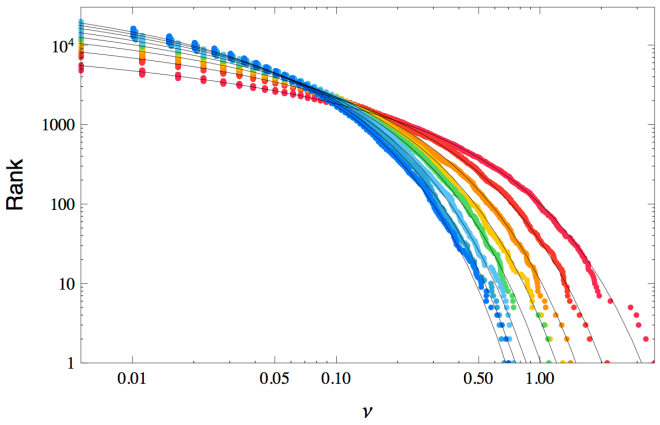


Figure 9: Rank-ν plots for all mutations that enter after $\sim$ 60 generations from simulations with $N_b = 10^7$, $U_b = 10^{-4}$, $s = 0.025$ in which cells are Poisson sampled with mean $2(1+s)$ for T generations then Poisson sampled with mean 2^{-T} on the Tth generation. Mean growth rates per generation are therefore $1+s$. ν is defined as $n/(e^{st}/s)$. The colors are different T: from 1-8 (red-blue). Although mean growth rates are the same, the distribution of sizes of mutants changes. The expected distribution of ν for constant N is $e^{-c\nu}/(c\nu\Gamma(NU))d\nu$. The black theory curves show the rank-ν plots predicted by this expression if we rescale $c \to c/T$ (or equivalently rescale $N \to NT$) as in the text. The simulation only allows single mutants.

Effective population size. In the analysis that follows it is sometimes necessary to use a population size in calculations. The most common is in the definition of the establishment time:

$$n(t) = (c/s)\exp\left(s(t-\tau_{est})\right) \tag{20}$$

To infer a τ_{est} we therefore need to relate frequencies (what we actually measure) to number. To do this we use the effective population size, $N_e = TN_b$ (for the population) or $n_e = Tn_b$ (for an individual lineage). We use this population size rather than the bottleneck size because it is this "effective" population size that correctly predicts the rate at which mutations enter the population which is the relevant property for the statistics of establishment times.

4.5 Multiple mutations and clonal interference within a lineage

Within a small lineage of n_0 cells it is possible that further beneficial mutations enter after the first one has established. This can occur in two distinct ways:

1. *Clonal interference within lineage.* If the population of n_0 cells or the mutation rate U_b are large enough, other independent single beneficial mutations will occur inside the lineage before the first has reached a size of $\sim n_0$ (clonal interference). This is undesirable because the expansion of the lineage will be driven by multiple independent beneficial mutations.

   Comparing Eqn. (5) with Eqn. (7) clonal interference effects within a lineage are important when $n_0U_b \gtrsim 1$. For $n_0U_b \ll 1$ beneficial mutations are rare so that the total number of cells with the beneficial mutation is dominated by the first one to establish. However, for $n_0U_b \gg 1$ the first n_0U_b mutations contribute significantly to the overall expansion and "interfere". To avoid clonal interference within a lineage and to be sure that any expansion of a lineage we measure is due to a *single beneficial mutation* lineage sizes must be small enough that $n_0U_b \ll 1$. What is the relevant U_b? In principle one should use the total beneficial mutation rate for U_b to completely avoid clonal interference within a lineage, however, in practice it takes at least of order $1/s$ generations for a mutation to reach establishment. Since we confine most of our analyses to the first $\sim$ 150 generations of evolution, mutations with $s \lesssim 0.01$ are largely irrelevant. This is discussed in more detail in Section 9.1. For

the time being we consider a reasonable range of $U_b \in \left[10^{-6}, 10^{-4}\right]$ which means that lineage sizes must be smaller than

$$n_0 \lesssim \begin{cases} 10^6 & \text{for } U_b = 10^{-6} \\ 10^4 & \text{for } U_b = 10^{-4} \end{cases} \tag{21}$$

Because multiple mutants involve how likely it is that mutations establish, the relevant lineage size we should use for comparison is the "effective" lineage size $n_e = n_b T$ which in our experiment is $\sim 10^3$ cells. The product is therefore $n_e U_b \sim 0.001 - 0.1$. Lineages are therefore well within the regime where expansion of the lineage is driven by a single beneficial mutation.

2. *Double-mutants.* Further beneficial mutations can also occur by one of the descendants of the first beneficial mutation acquiring a second beneficial mutation, creating a double-mutant. If the single mutant grows as $(c/s)e^{s(t-\tau_1)}$, the probability of a double mutant entering the population and establishing will be appreciable after a time τ_2 at which

$$U_b \times (s/c) \times \int_0^{\tau_2} (c/s) e^{s(t-\tau)} dt \sim 1 \tag{22}$$

Giving

$$\tau_2 \sim \tau_1 + (1/s)\log(s/U_b) \tag{23}$$

For $s \sim 0.05$ and $U_b \in \left[10^{-6}, 10^{-4}\right]$ (consistent with out later inferences) double mutants will typically emerge $\sim 180 - 220$ generations after single mutants. Since our experiment is confined to the first ~ 150 generations the number of double mutants is therefore expected to be small. We can quantify how many of the observed adaptive lineages are likely to be double mutants in the following way. Consider a lineage that has accumulated a beneficial mutation which is growing as

$$n_1(t) = \frac{c e^{s_1(t-\tau_1)}}{s_1} \tag{24}$$

In order for the observed increase of the lineage to be due to a double mutant two things must be true: (i) a double mutant must have established and (ii) it must have grown to (at least) as large a size as the first mutant (if not we would still largely measure the effect of the first mutant). If the double mutant then grows as

$$n_2(t) = \frac{c e^{s_2(t-\tau_2)}}{s_2} \tag{25}$$

The condition that the second mutant be larger than the first means that

$$\tau_2 < \frac{t + \tau_1}{(s_2/s_1)} = \frac{t + \tau_1}{2} \tag{26}$$

where we have assumed that double mutants have a fitness $s_2 = 2s_1$. Considering the typical establishment time of early single mutants in the range of $-100 < \tau_1 < -50$, this means the establishment time of the double mutant must be close to zero or even negative. The probability that a double mutant establishes before this time is

$$P\left(\tau_2 < \frac{t + \tau_1}{2}\right) = \frac{U_b s_2}{s_1^2} e^{-s_1 \tau_1} \approx 10^{-2} \tag{27}$$

(putting in typical values of $U_b \approx 10^{-5}$, $\tau_1 \approx -100$ and $s \approx 0.04$). Of the $\sim 25,000$ observed mutations it is unlikely that more than ~ 500 are double mutants. Given we observe $\sim 3,000 - 5,000$ in the high fitness range ($s > 5\%$) we conclude that most of these are still single mutants.

Although the fraction of adaptive mutant lineages that have acquired a second mutation will be small, the total number of double mutant cells can become large before the single mutants collectively take over the population. For example if the mutations all had the same fitness increment, in the deterministic approximation the fraction of the population that are double mutants is

$$f_{double} \approx \frac{1}{2} U_b^2 / s^2 e^{2st} \approx \frac{1}{2} f_{single}^2 \ . \tag{28}$$

Although this deterministic approximation breaks down because $NU_b^2/s \ll 1$, one does still expect that the double mutants will be a substantial fraction of the population soon after the single mutants begin to dominate, though in our case this occurs at $t > 150$ generations.

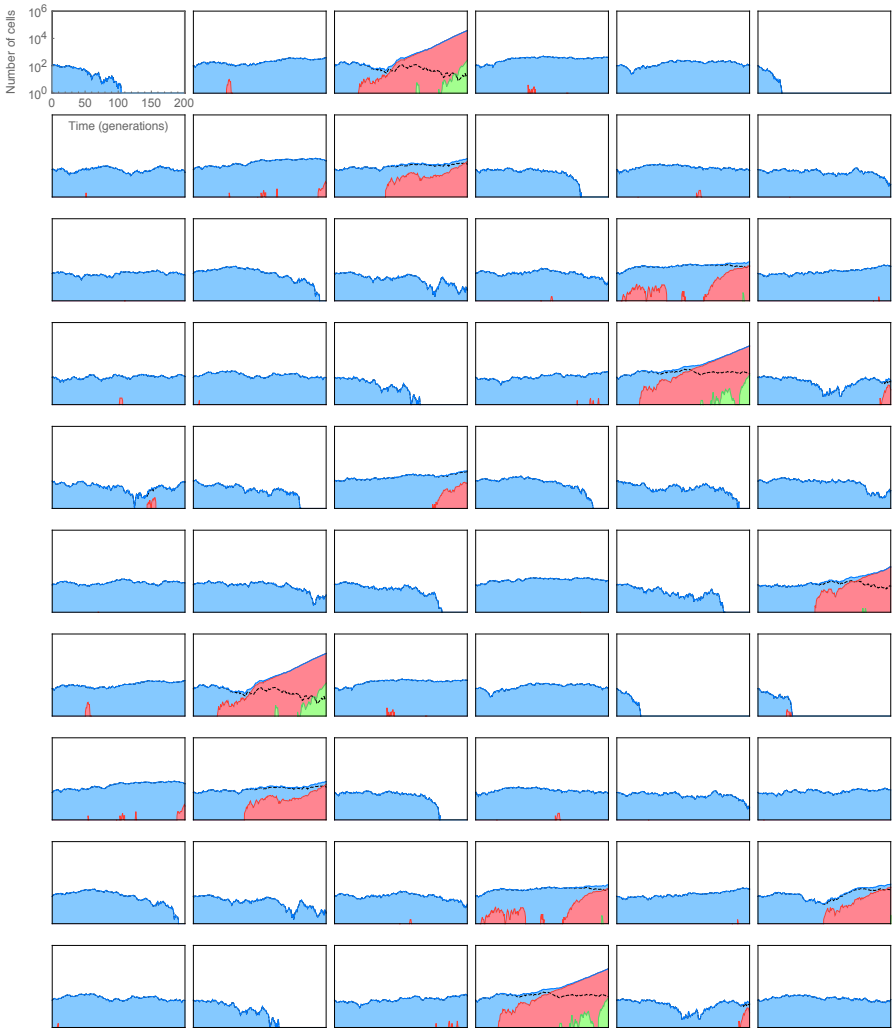


Figure 10: Simulated trajectories of initially neutral lineages (blue, $n_0 = 100$ cells) that feed beneficial mutations (red) with fitness $s = 5\%$ at a rate $U_b = 10^{-5}$. Beneficial mutations can themselves accumulate a further $s = 5\%$ mutation becoming a double mutant with a total fitness advantage of 10% (green). Most lineages either never accumulate a beneficial mutation or do so, only for it to drift out. In the cases where the beneficial mutation establishes and grows exponentially, there is a window in which the expansion of the lineage is dominated by the single mutant, since any double mutant that enters has not had long enough to overtake the single mutant.

5 Noise model

To assign likelihoods to barcodes containing beneficial mutations we need an error model describing how likely trajectories are in the absence of any mutation. This depends on variations in abundance introduced through various processes involved in measurement including: sequencing, extraction and amplification as well as the stochastic fluctuations though multiple growth-dilution cycles. In the following sections we outline how to quantify each of these.

5.1 Sequencing noise

The simplest approximation is that sequencing introduces binomial (or Poisson) sampling errors in frequencies with variance $Rf(1-f) \approx Rf$. The difference between binomial and Poisson are negligible since $f \ll 1$. To test this hypothesis we consider f (the true frequency of the barcode in the DNA sent to the sequencer) and ask for the statistics of the joint distribution $P(r_1, r_2|R_1, R_2)$ from two *independent* sequencing runs with R_1 and R_2 total reads *but the same DNA prep* (so that f is the same for both, Figure 11). Under some prior probability density $\rho(f)$ we have

$$P(r_1, r_2) = \int_0^1 P(r_1|f)P(r_2|f)\rho(f)df \tag{29}$$

Substituting in the standard form for the Poisson distribution we find

$$P(r_1, r_2) = \frac{R_1^{r_1}}{r_1!}\frac{R_2^{r_2}}{r_2!}\int_0^1 f^r \exp(-fR)\rho(f)df \tag{30}$$

where $r = r_1 + r_2$ and $R = R_1 + R_2$. We notice that multiplying and dividing by $r!/R^r$ we can write this as

$$P(r_1|r)P(r) = \binom{r}{r_1}\left(\frac{R_1}{R}\right)^{r_1}\left(\frac{R_2}{R}\right)^{r-r_1}\underbrace{\int_0^1 \frac{(fR)^r \exp(-fR)}{r!}\rho(f)df}_{=P(r)} \tag{31}$$

For the Poisson distribution (and for the binomial), the distribution of r_1 conditioned on the sum $r = r_1 + r_2$ is binomial, with probability $p = R_1/R$ *independent of the prior*. Taking logs, using Stirling's approximation and keeping only terms that have an r_1 dependence we have

$$\ln P(r_1|r) = r_1 \ln\left(\frac{p}{r_1}\right) + (r - r_1)\ln\left(\frac{1-p}{r-r_1}\right) \tag{32}$$

Figure 12 compares this expression from Eq. 32 to experimental data from independent sequencing runs *from the same DNA prep*. The plots show that errors introduced by a finite number of reads at the sequencer are very well approximated by a Poisson sampling.

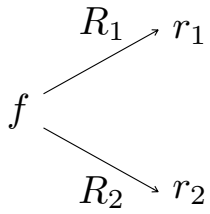


Figure 11: A single sample has DNA extracted and amplified but sent to the sequencer on two independent runs. A barcode present at frequency f in the common DNA pool is then sampled R_1 or R_2 times total with mean $R_1 f$ or $R_2 f$

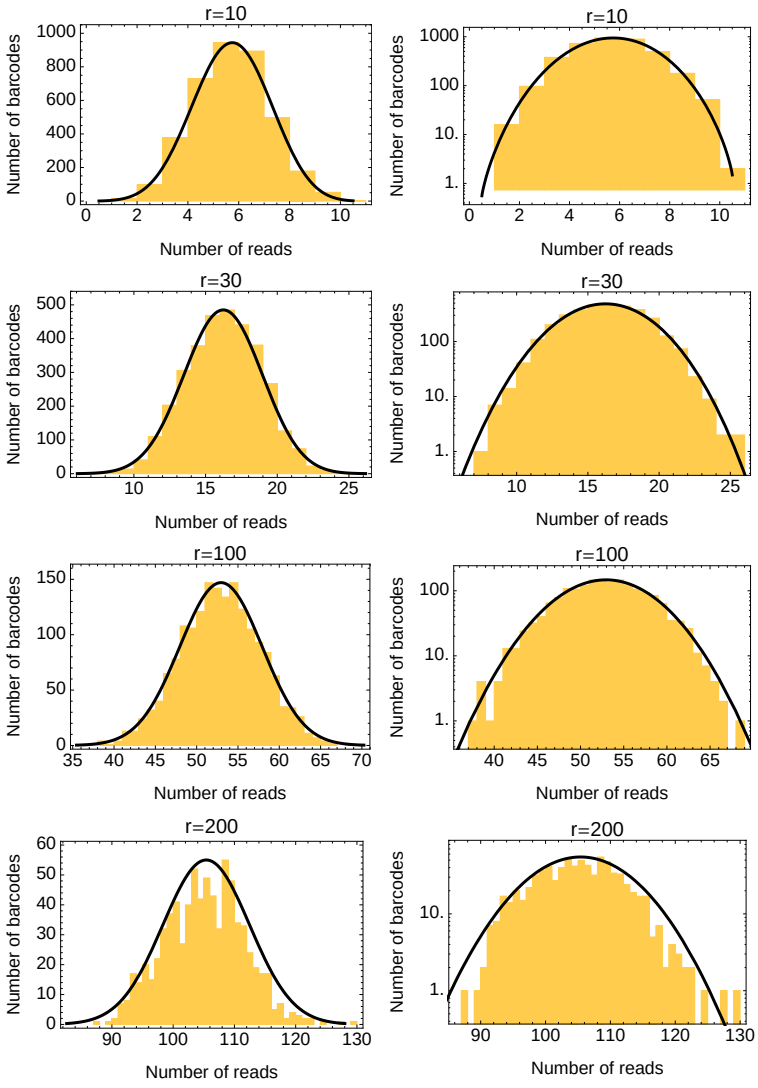


Figure 12: Distributions of numbers of reads r_1 read on a second sequencing run for barcodes that were read a total of $r = r_1 + r_2$ times across both sequencing runs. The total read depth on the first run is $R_1 = 2.66 \times 10^7$ while on the second it is $R_2 = 2.95 \times 10^7$. Theoretical curves are produced assuming Poisson sampling with mean r_1/R.

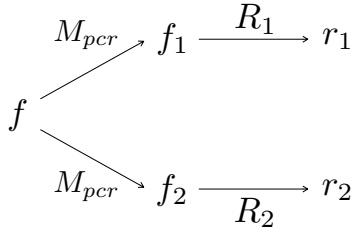


Figure 13: A single pool of cells is independently prepped to extract and amplify DNA (modeled as a sampling to M_{pcr}, which is a free parameter) then sent to the sequencer on two independent runs (modeled as samples of size R_1 or R_2, which are measured quantities)

5.2 Sequencing + Amplification noise

If the DNA is extracted and amplified independently in addition to being sequenced independently, there is additional noise due to the extraction / amplification (Figure 13). The distribution of reads r that result from a barcode present at frequency f in the population that goes through a DNA extraction and amplification followed by sequencing is modeled using a form similar to that derived in Section 13.3 for the distribution from stochastic drift:

$$P(r|a) \approx \sqrt{\frac{a^{1/2}}{4\pi b r^{3/2}}} \exp\left[-\frac{(\sqrt{r}-\sqrt{a})^2}{b}\right] \tag{33}$$

where r is the observed number of reads, $a = Rf$ is the mean number of reads expected and $b = 1/2(1 + R_1/M_{pcr})$ controls the variance ($\sigma^2 = 2ab = Rf(1 + R/M_{pcr})$). We use M_{pcr} to control the additional noise introduced by the extraction and amplification process, treated as an additional sampling of size M_{pcr}, the magnitude of which we can tune. The additional noise term can then be understood by realizing that, for a barcode at frequency f, the sampling of M_{pcr} molecules introduces a variance in frequency of magnitude f/M_{pcr}. When this is read at the sequencer R_1 times, this translates to a variance of magnitude $R_1^2 f/M_{pcr}$ hence the form of the noise term term. Setting $M_{pcr} \to \infty$ recovers the correct mean and variance from sequencing alone that was worked out in the previous section.

Why do we use this form for the distribution, instead of a Gaussian?. While our distributions are controlled only by a mean (Rf) and a variance ($Rf(1+R/M_{pcr})$) fitting a Gaussian distribution from these parameters would be the wrong thing to do. Our data strongly suggests (see Figure 15) that far above the mean the probability decays closer to exponentially ($\exp(-r/b)$) — as is expected both for Poisson sampling and for neutral drift of the cell populations as analyzed in section 13.3 — than as a Gaussian ($\exp(-r^2/ab)$). This makes a significant difference far in the tail of the distribution, which is particularly important in our case since we need to distinguish rare neutral events from beneficial ones.

Using the distribution in Eqn. 33 we can calculate the joint distribution for reads r_1 and r_2 from two independent amplifications and sequencing runs:

$$P(r_1, r_2) \approx \int P(r_1|f)P(r_2|f)\rho(f)df \tag{34}$$

Although we do not know f for any one barcode, its distribution is very close to being exponential with a decay length of $\bar{f} \approx 2 \times 10^{-6}$ (Figure 14), which we use as our prior. This decay length corresponds to a "typical" initial lineage size of ~ 120 cells at bottleneck. We can therefore compute $P(r_1, r_2)$ from the integral in Eqn. 34 setting the parameters

$$a_1 = R_1 f \qquad b_1 = (1/2)(1 + R_1/M_{pcr}) \tag{35}$$

$$a_2 = R_2 f \qquad b_2 = (1/2)(1 + R_2/M_{pcr}). \tag{36}$$

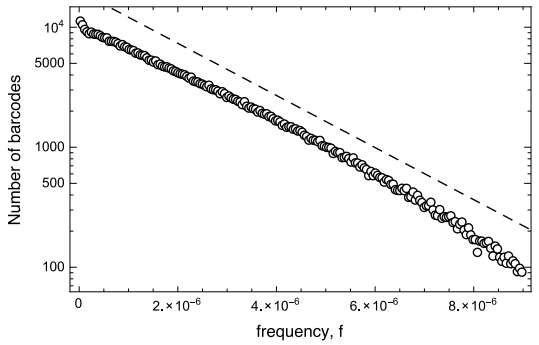


Figure 14: The initial distribution of barcode frequencies by pooling all ~ 222 million t_0 reads is approximately exponential over a large range with decay length $\bar{f} \approx 2 \times 10^{-6}$.

The variance in frequency has an additional contribution with magnitude $1/M_{pcr}$ to account for the extraction / amplification noise. We can use this as a free parameter and fit it to see how much additional noise is introduced by DNA extraction and amplification. These best fits are shown in Figure 15 (black line) compared to the case where there is only sequencing noise present (dashed line). The additional noise due to DNA amplification is equivalent to an additional sampling of $M_{pcr} \approx 7.3 \times 10^7$. Comparing this to the typical sampling introduced by reads at the sequencer (where $R \approx 3 - 5 \times 10^7$) we see that DNA amplification typically introduces about half as much variance in read numbers compared to the sequencing. Crucially, the DNA amplification does not appear to result in a substantially longer tail as could potentially have occurred from variable amplification biases.

Total errors in frequency measurement. The above analysis of the two factors contributing to noise in barcode frequencies inferred at a given time point show that errors in frequencies are

$$\frac{\delta f}{f} \approx \frac{1}{\sqrt{f}} \left(\frac{1}{R} + \frac{1}{n} \right)^{1/2} \approx \frac{2 \times 10^{-4}}{\sqrt{f}} \tag{37}$$

Barcodes are initially at a frequency of $\sim 10^{-6}$ meaning errors in frequency are $\sim 20\%$. However if lineages become adaptive and increase in frequency (adaptive linages typically have $f \gtrsim 10^{-5} - 10^{-4}$) errors fall to the few percent range.

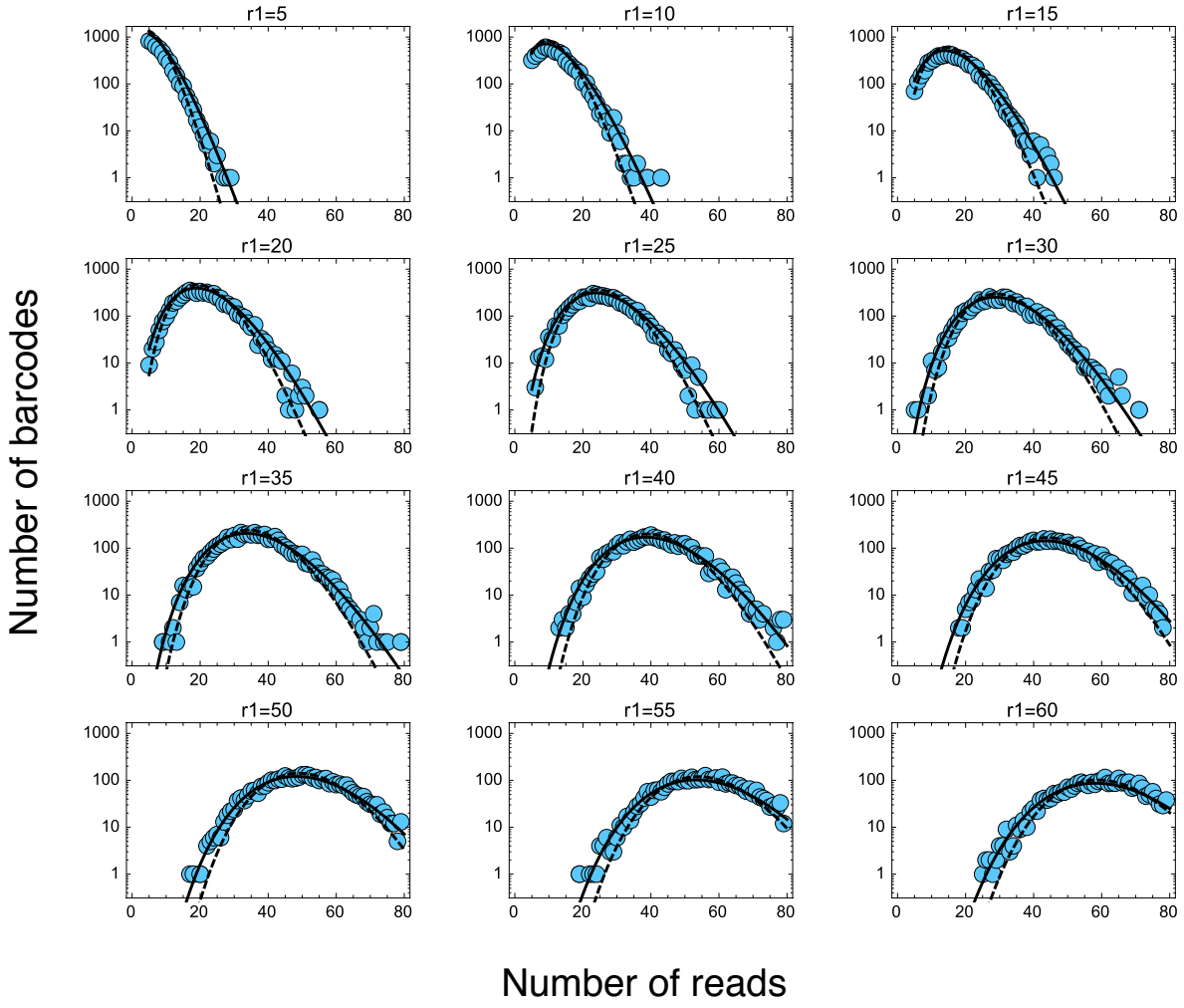


Figure 15: Slices through the joint distribution of reads $P(r_1, r_2)$ for two samples that have been prepped to extract and amplify DNA independently and then and sequenced independently. The measured distribution across two independently prepped samples (solid curve) is clearly wider than would be predicted by sequencing noise alone (dashed curve). The additional variance amounts to an effective sampling of size 7.3×10^7 which means the DNA extraction / amplification adds approximately half the variance typically introduced by sequencing (which typically has a sample size of $R \sim 3 \times 10^7$).

5.3 Sequencing + Amplification + Growth noise

If the population is evolved through some number, q, of growth-bottleneck cycles before r_2 is measured there is additional "noise" due to the variance introduced both during the bottleneck and during the growth phase (Figure 16). Since we directly infer the bottleneck size, N_b, any variance above that expected from this is due to variations in the cellular growth rates, from which c (half the effective variance in offspring number per generation) can be estimated. As highlighted earlier, because the population comes out of stationary phase at the beginning of the cycle, variations in lag time and variations in the division rate in the next few divisions are likely to dominate the variations. The additional variance measured over q cycles is $2c(q/N_b)$ where q is the number of cycles, $N_b \approx 6 \times 10^7$ is the (known) bottleneck size and $2c$ a free parameter that is the variance across the cycle from the bottleneck and growth. The parameter M_{pcr} from the previous step is now fixed. Again the distribution remains of the same form as in Eqn. 33 but the

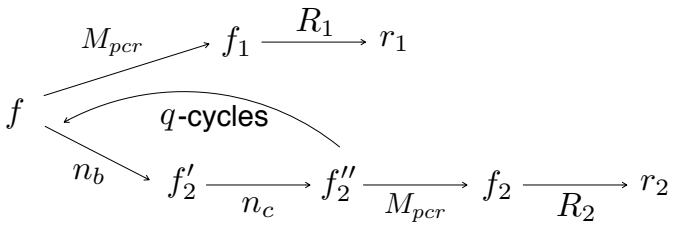


Figure 16: Model of additional variance introduced by q growth bottleneck cycled on top of the noise introduced by DNA preps and sequencing.

parameters are updated to include the additional variance:

$$a_1 = R_1 f \qquad b_1 = (1/2)\left(1 + R_1\left(\frac{1}{M_{pcr}} + 2c\frac{q}{N_b}\right)\right) \tag{38}$$

$$a_2 = R_2 f \qquad b_2 = (1/2)\left(1 + R_1\left(\frac{1}{M_{pcr}} + 2c\frac{q}{N_b}\right)\right) \tag{39}$$

The variance in frequency therefore has an additional contribution of size $2cq/N_b$ from the q bottleneck-growth cycles. Since N_b is fixed c is a free parameter which we can fit across varying number of cycles q to see how much additional noise is introduced by growth of the cells (Figure 17).

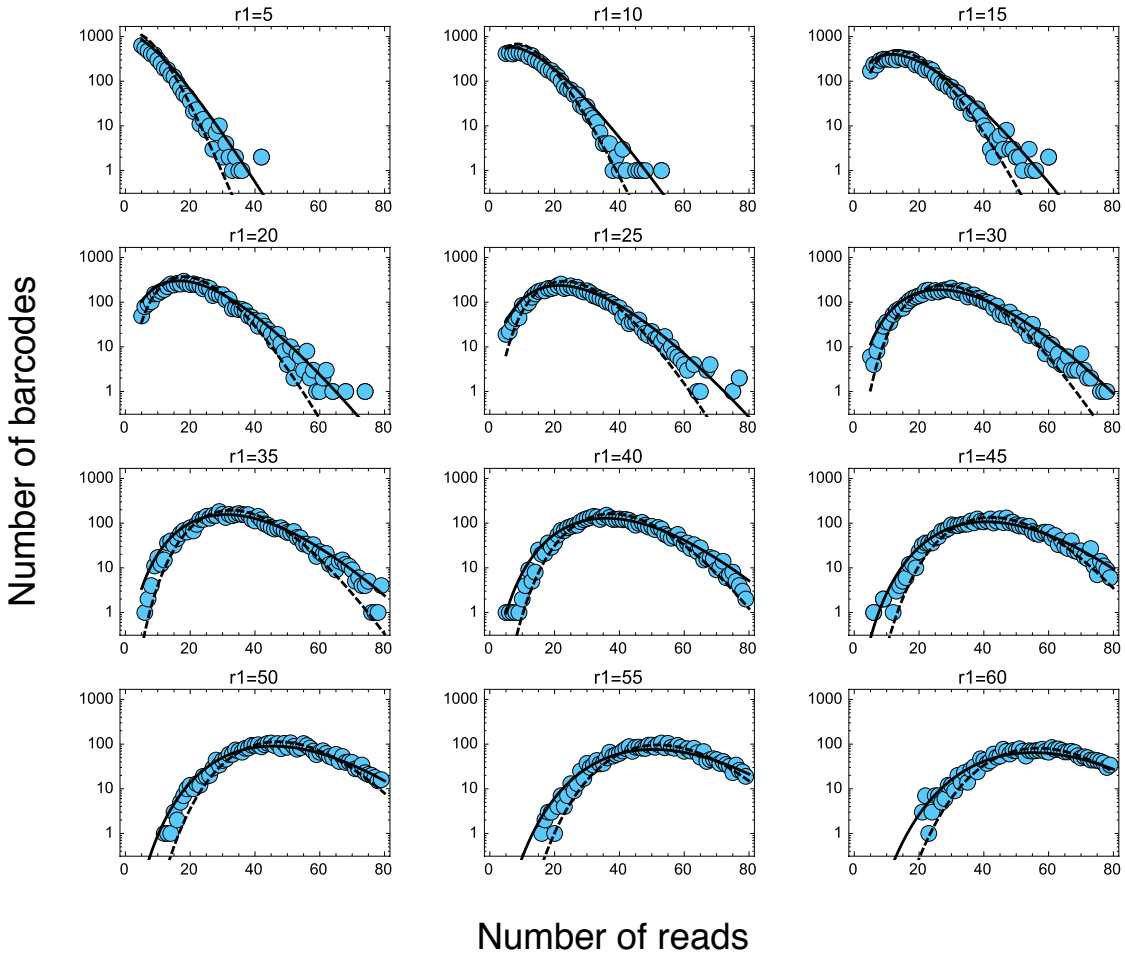


Figure 17: Slices though the joint distribution between samples from $t = 0$ and $t = 16$ i.e. $(q = 2)$. The variance explained by sequencing noise, amplification noise and bottleneck sampling (dashed curve) is not enough to explain the data. Additional noise due to the variation in offspring number during growth (solid curve) must be included. The effective single cycle best fit yields a variance of $2c \approx 3.5$

5.4 The effective offspring distribution though a cycle

The previous section can now be used to estimate the effective offspring distribution for a single cell across the growth-bottleneck cycle. The total variance in offspring number across the cycle is simply the variance in frequency multiplied by the bottleneck size N_b which gives

$$var(\text{number of offspring across cycle}) = 2c \approx 3.5 \tag{40}$$

In the case of Poisson sampling each generation the value of $c = 0.5$ hence the variations introduced are substantially more than this, probably due to variations in lag time. We verify inferred value of c is consistent across multiple growth-bottleneck cycles by plotting it across multiple cycles in Figure 18.

What about higher moments? So far we have classified growth through the cycle using a mean (= 1 for neutral cells or $1 + s$ for beneficial cells) and a variance (≈ 3.5) in the number of offspring from a single cell at the bottleneck. Our data suggest that the form we propose for the distribution of offspring number though a cycle — that it falls off exponentially rather than as a Gaussian at large numbers — is sufficient to capture the probability of large jumps in frequency. We find no evidence of a longer tail in frequency jumps between time points. While it is possible that the "true" distribution of offspring number is slightly more skewed than proposed here, across many time points these higher moments are less and less important

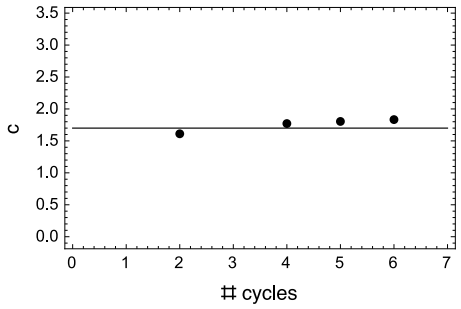


Figure 18: The fitted parameter c (half the effective variance in offspring number per cell per generation) during growth is consistent when measured across many cycles of increasing lengths. Fitting the value of c for $q > 6$ cycles gives values of $c \sim 7$, however this is expected since across this many generations ($t > 64$) adaptive mutations come in that contribute anomalously to variance estimates.

(analogous to, although not the same as, the central limit theorem becoming better for more samples) and the exponential form for the tail appears to be sufficient.

6 Inference of mean fitness

Once beneficial mutations expand and reach appreciable frequencies they change the mean fitness of the population. We would like to infer how this mean fitness, $\bar{x}(t)$, increases since it is a measure of how rapidly the population is adapting and because it determines how rapidly mutations enter later in the evolution.

One method to infer the mean fitness is to monitor how rapidly neutral lineages decline in frequency. If the mean fitness is $\bar{x}(t)$, the relative change in frequency of a neutral lineage between $t - \Delta t/2 \to t + \Delta t/2$ is

$$\delta f/f = \exp\left(-\int_{t-\Delta t/2}^{t+\Delta t/2} \bar{x}(t)dt\right) \approx \exp\left(-\bar{x}\left(t\right)\Delta t\right) \tag{41}$$

Where the last approximation assumes the timescale over which the mean fitness changes is long relative to Δt and so that $\bar{x}(t)$ can be assumed constant. Measuring the rate of decline of neutral lineages therefore gives one an estimate of the mean fitness. This is of a similar flavor to the methods used to estimate mean fitness in [14].

Another way to measure mean fitness is to identify a putative set of all adaptive lineages, and use the fitness estimate for each of these to explicitly calculate the mean fitness using

$$\bar{x}(t) \approx \sum_j s_j f_j(t) \tag{42}$$

where j enumerates the set of adaptive lineages. (Note that the estimate of s_j itself must use information about the increasing mean fitness as discussed in Section 7).

These two way of estimating the mean fitness should agree with one another, which can be used as a self-consistency check. Our approach will be to first measure the mean fitness by monitoring how rapidly neutral barcodes decline in frequency. We will then use this to pick out lineages that are likely adaptive and check that these give an estimate of the fitness that is self consistent with the fitness we inferred from the neutrals.

6.1 Using low abundance lineages as neutral markers

The vast majority of low frequency lineages (present in $\sim 20-50$ cells at the bottleneck) will not accumulate a beneficial mutation before ~ 1000 generations. If beneficial mutation rates are $\sim U_b \sim 10^{-5}$ and selection coefficients about $s \sim 0.05$ (both of which broadly agree with our later inferences) then a lineage that initially has an effective size of n_e cells will likely accumulate a beneficial mutation after a time t when

$$n_e U_b (s/c) t \approx 1 \tag{43}$$

Giving a typical waiting type for the accumulation of a mutation of

$$t \sim \frac{1}{n_e U_b (s/c)} \approx 2000 \tag{44}$$

generations for a lineage with ~ 20 cells at the bottleneck. By ~ 150 generations (the time to which we restrict our analysis) only $\sim 7\%$ of low frequency lineages will have accumulated a beneficial mutation, hence they are a good candidate set of neutral lineages. Even if a small fraction do acquire beneficial mutations that establish, they will not remain at low frequencies for longer than a few times $1/s$ generations and hence will not greatly affect results.

To infer the mean fitness from the candidate set of neutral lineages we:

- Form a list of all barcodes that were read exactly r_t times at time point t. We restrict ourselves to $20 \leq r_t \leq 40$ reads (see below for explanation). Given typical read coverage of $\sim 3 \times 10^7$ per time point and a bottleneck population size of $\sim 7 \times 10^7$ those barcodes read $r \approx 20$ times correspond to ≈ 40 cells at the bottleneck.

- Plot the distribution of these reads at the subsequent time point $t+T$, where T is the number of generations to the next sequenced time point. This is usually 8 generations though sometimes 16 or 24 generations elapse between sequenced time points.

- Use the predicted distribution of reads r_{t+T} conditioned on r_t (worked out in Section 5) that includes the mean reduction in barcode abundance expected due to competition against a mean fitness ($e^{-T\bar{x}}$)

$$P(r_{t+T}|r_t) \approx \sqrt{\frac{(r_t e^{-T\bar{x}})^{1/2}}{4\pi\kappa r_{t+T}^{3/2}}} \exp\left[-\frac{\left(\sqrt{r_{t+T}} - \sqrt{r_t e^{-T\bar{x}}}\right)^2}{\kappa}\right] \tag{45}$$

  to obtain a best fit value for the mean population fitness, $\bar{x}$, and the noise parameter κ between time point t and $t+T$ for the given r_t. Here κ parametrizes the total fluctuations and noise across the cycle, including effects from sequencing, amplification and the growth-bottleneck cycle. It is approximately equal to $1/2 + R/M_{pcr} + 2cR/N_b \approx 3$ and is largely independent of the frequency of the barcode. The best fit is defined as the pair $(\bar{x}, \kappa)$ that minimizes `distance`, which is defined as the summed square differences between the predicted distribution and the measured distribution:

$$\texttt{distance} = \sum_{j}^{2r_t} (\text{Measured } \# \text{ at } r_j - \text{Predicted } \# \text{ at } r_j)^2 \tag{46}$$

  Where the predicted reads comes from Eqn. 45. For each r_t and each time point this yields a best-fit pair $(\bar{x}, \kappa)$.

- Repeat this for $20 \leq r_t \leq 40$. We focus on these read numbers because these lineages are unlikely to have accumulated beneficial mutations. We avoid using barcodes read < 20 times because for low integer read numbers the approximate form for the distribution of reads Eqn. 45 does not quite capture the distribution correctly at very low read numbers (see Section 5).

- Estimate the mean fitness at each time point taking the mean of the estimates over the different r_t.

$$\bar{x}_t = \frac{1}{21}\left(\bar{x}_{20} + \bar{x}_{21} + ...\bar{x}_{40}\right) \tag{47}$$

  We also determine the mean best-fit κ at each time point via

$$\kappa_t = \frac{1}{21}\left(\kappa_{20} + \kappa_{21} + ...\kappa_{40}\right) \tag{48}$$

- Repeat for each time point to obtain an estimate for $\bar{x}$ between each subsequent time point. This method of calculating the mean fitness between two subsequent time points performs well at inferring the mean fitness of the population when compared to simulations in which the mean fitness is known (see Section 12).

Figure 19 shows the measured distributions of $r_{t+T}|r_t$ at each time point (rows) for various r_t (columns). The solid black curve is the expression from Eqn. 45 with the best fit $(\kappa, \bar{x})$ pair. The dashed black curve is the predicted distribution Eqn. 45 with $\bar{x} = 0$. Initially there is little difference between the solid and dashed curves meaning that the mean fitness remains $\bar{x} \approx 0$. However, at times ($t > 64$) the observed distribution of reads r_{t+T} (yellow bars) is clearly skewed to lower read numbers compared to what would be expected if there were no change in mean fitness. This skew becomes pronounced at later times where a non-zero mean fitness is clearly needed to explain the change in frequency of the barcodes.

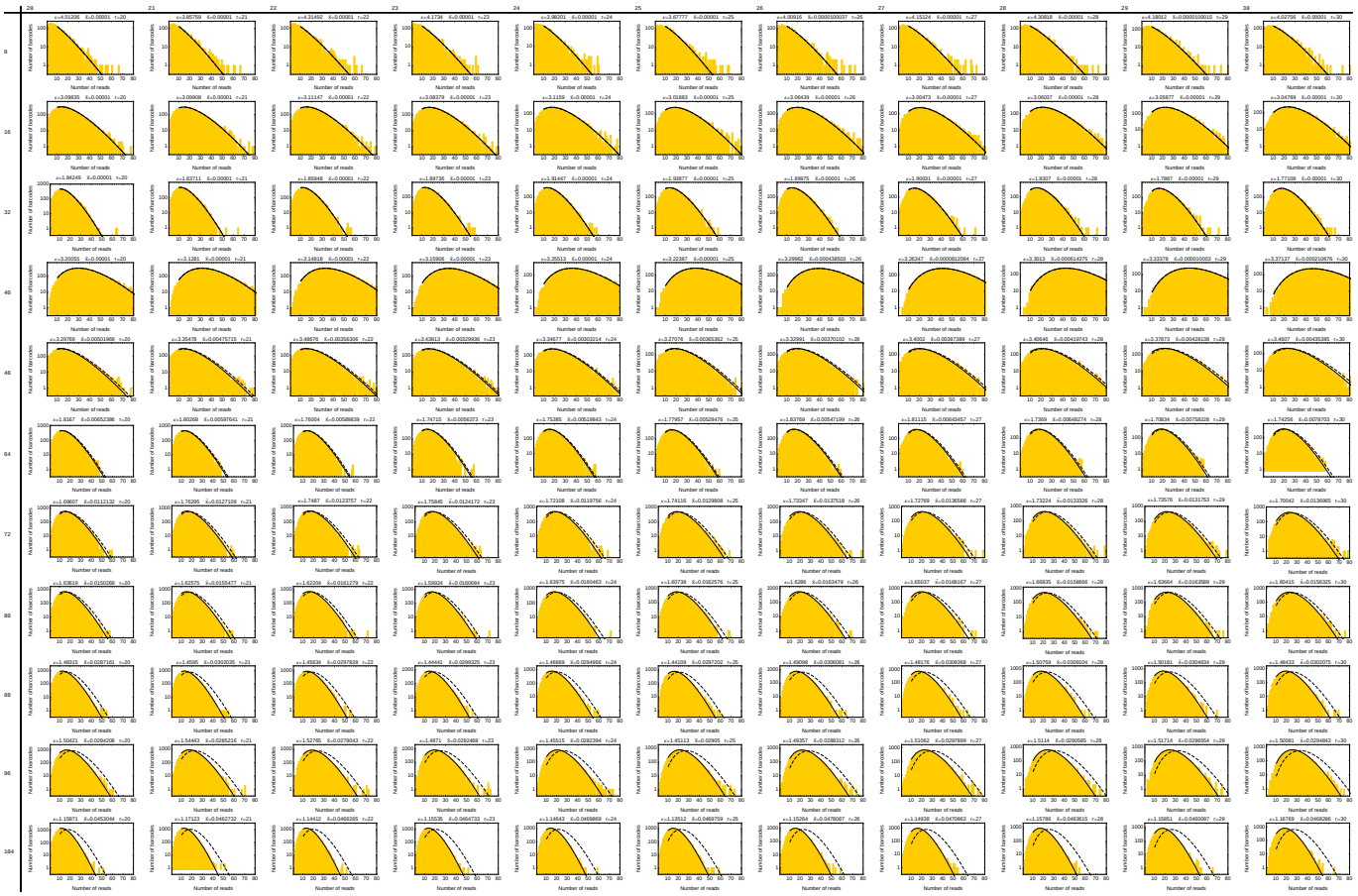


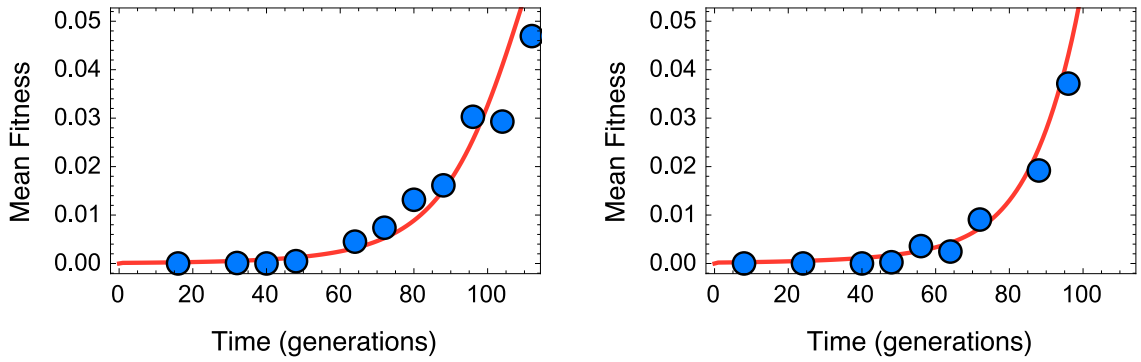
Figure 19: Inferring the mean fitness from neutral lineages. All lineages read r_t times at time point t are identified. The distribution of reads of these lineages at a subsequent time point r_{t+T} is plotted (yellow bars). We fit the distribution with its expected theoretical form from Eqn. 45 for all $20 < r_t < 40$ (columns of the plot, shown up to 30) for all neighboring time points (rows of the plot). All data from replicate E1. The solid black curve is the predicted distribution of reads from Eqn. 45 with the best-fit values values for $(\kappa, \bar{x})$ (shown above each plot). The dashed curve is the predicted distribution of reads if mean fitness $\bar{x} = 0$. At later times the true distribution of reads is skewed to lower read numbers than predicted by assuming $\bar{x} = 0$ because these neutral lineages are being "squeezed out" by the increasing mean fitness.

6.2 Using a set of adaptive lineages to infer mean fitness

Once a putative set of adaptive lineages has been identified and estimates for the fitness effect of the mutation within the lineage s_j, as well as the abundance f_j of the lineage are available one can explicitly calculate the mean fitness of population via

$$\bar{x} = \sum_j s_j f_j \tag{49}$$

where j enumerates the set of putatively adaptive lineages (outlined in the following section). The inference of the fitness s_j for each adaptive lineage itself depends on the mean fitness that we infer using the low-abundance neutral marker approach described above (as outlined in Section 7). For consistency, and to be confident that we are capturing the majority of lineages that contribute to the mean fitness, the value of $\bar{x}$ inferred using the adaptive lineages should agree with the mean fitness inferred by monitoring how rapidly the neutral barcodes decline in frequency. The comparison between these two ways of inferring the mean fitness for E1 and E2 are shown in Figure 20a and 20b respectively. Given the good agreement and the



(a) E1 **(b)** E2

Figure 20: Plots comparing the mean fitness inferred from the decline of neutral lineages (blue circles) to the mean fitness inferred from the expansion of beneficial lineages (red line) for both replicates E1 and E2. The consistency between these two methods confirms that we are successfully identifying the majority of lineages that contribute to the increasing mean fitness.

uncertainties in each, we estimate that the accuracy of the inferred mean fitness is better than $\approx 0.001\%$ at early times (when it has not increased much above $s = 2\%$) and around 0.5% at $t \approx 100$

6.3 Using the local log-gradient of a trajectory and its measured abundance to infer mean fitness

The method described in the previous section used the *inferred* values of (s, τ) for each barcode to estimate the number of adaptive cells there are in each lineage and thus calculate the mean fitness. There is a more direct way of doing this that does not rely on inferences from the entire trajectory. The abundance of each lineage is measured and the fitness of the cells within that lineage can be estimated *locally* by measuring the log-gradient of that trajectory (which in the limit of the lineage being dominated by a single expanding mutation, is a good estimator of its fitness).

The mean fitness measured using this method is clearly not independent from that of the previous section. However it gives an estimate that is local: relying only on the measured abundance of the time point in question and its immediate neighbor. To infer the mean fitness using this approach we:

1. Collect the set of putatively adaptive lineages, $\mathcal{AL}$ by selecting all those for whom there is evidence of them being adaptive ($\ln$ (maximum posterior of beneficial) $-\ln$ (posterior being neutral) > 0)

2. Initialize the mean fitness to zero: Mean Fitness$[i = 0] = 0$

3. Estimate the local lineage fitness

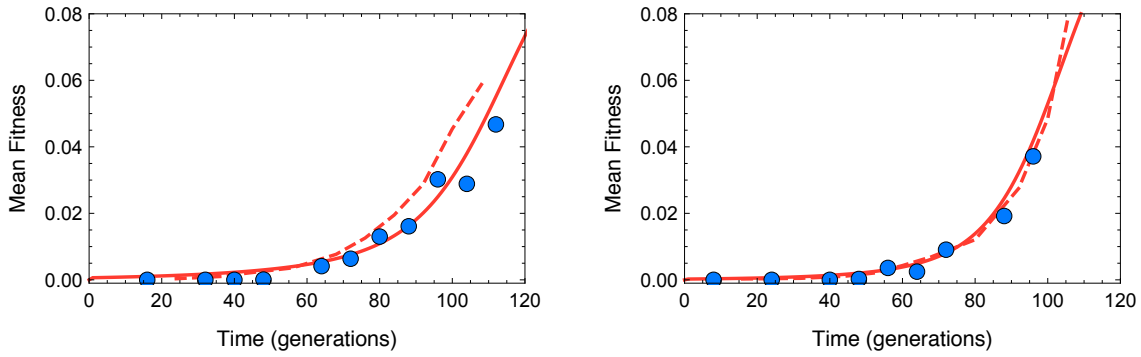
$$\text{Fitness}\,[i+1] = \text{Mean Fitness}\,[i] + \frac{\log(\text{Abundance}\,[i+1]) - \log(\text{Abundance}\,[i])}{\text{Time}\,[i+1] - \text{Time}\,[i]} \tag{50}$$

4. Update the estimate of the mean fitness using the inferred local fitnesses and an estimate for the abundance of the beneficial mutants in each lineage:

$$\text{Mean Fitness}\,[i+1] = \sum_{\mathcal{AL}} (\text{Abundance}\,[i+1] - \text{Abundance}\,[i=0]) \times \text{Fitness}\,[i+1] \tag{51}$$

5. Repeat across all transfer time points i.

This locally inferred mean fitness is shown in 21. It broadly agrees with previous inferences though gives an estimate slightly above the two previous methods for the replicate E1.



(a) E1 **(b)** E2

Figure 21: Plots comparing the mean fitness inferred from the decline of neutral lineages (blue circles), the mean fitness inferred from the expansion of adaptive lineages (red line) from *inferred* values of (s, τ), and the mean fitness from locally measuring the log-gradient of each expanding adaptive lineage (dashed red line). Replicate E1 shows a slightly larger mean fitness via this method compared to the previous two methods. E2 agrees well for all three methods. It is possible that the discrepancy in E1 may account for some of the systematic differences in inferred fitnesses between replicates E1 and E2 described in Section 8.2. However we also note that at late times, the mean fitness is driven by a small number of fit mutations that are likely more sensitive to conditions compared to averages over a very large number of mutations, hence one may expect variability at later times.

6.4 Using the pre-existing mutation class to infer mean fitness

Another method to infer the population mean fitness that *is* almost entirely independent of the previous methods, uses the pre-existing lineages (Section 10) with fitness $s \approx 4\%$. In flavor this method is similar to the method of using neutral lineages in that it uses a set of lineages for whom we know the fitness. The pre-existing lineages likely contain mutations that occured prior to the separation of the two replicates and was sampled (and subsequently established) in both E1 and E2. We have many such "pre=existing" lineages with fitnesses between $0.03 < s < 0.05$. Here we group these lineages together and track the aggregate trajectory of all across both E1 and E2.

Because we know these lineages have a mutation that confers a benefit of $s \approx 0.04 \pm 0.005$, by tracking how the aggregate trajectory's gradient bends to become smaller than $s = 0.04$, we can estimate the mean fitness of the population. Concretely, the estimate is

$$\text{Mean fitness}\,[t] = 0.04 - \text{log-gradient}\,[t] \tag{52}$$

Figure 22 shows these aggregate trajectories compared to the expected $s = 0.04$ line. The fact that they track one another so closely until $t = 96$ demonstrates that the mean fitness of the population must be similar across the two replicates (consistent with our other inferences). However there is a dramatic departure between E1 and E2 for $t > 96$. This is consistent with the fact that at later times the mean fitness begins to behave stochastically and one expects there to be differences between the mean fitness of the replicates from the emergence of rare high fitness mutations.

The mean fitness inferred using this method is shown as the purple line in Figure 23 **(a)** and **(b)** where we have also shown the three previous measures of mean fitness for comparison. We estimated the magnitude of systematic errors by an upper and lower bounds that includes the mean fitness estimates from all four methods. These are shown as shaded regions in Figure 23 **(c)**. All four methods agree to within $\approx 1\%$.

We note that there is a caveat to using this method to infer the mean fitness which is that we have ignored the possibility of the accumulation of further beneficial mutations. Further beneficial mutations in the lineages that make up the aggregate purple lineage would increase its fitness and therefore would cause a *smaller* mean fitness estimate (because it would depart less from the 4% line). We see no evidence for this effect however which in itself could be tentative evidence that not many of these 4% lineages are acquiring further beneficial mutations in order to keep up with the moving mean fitness.

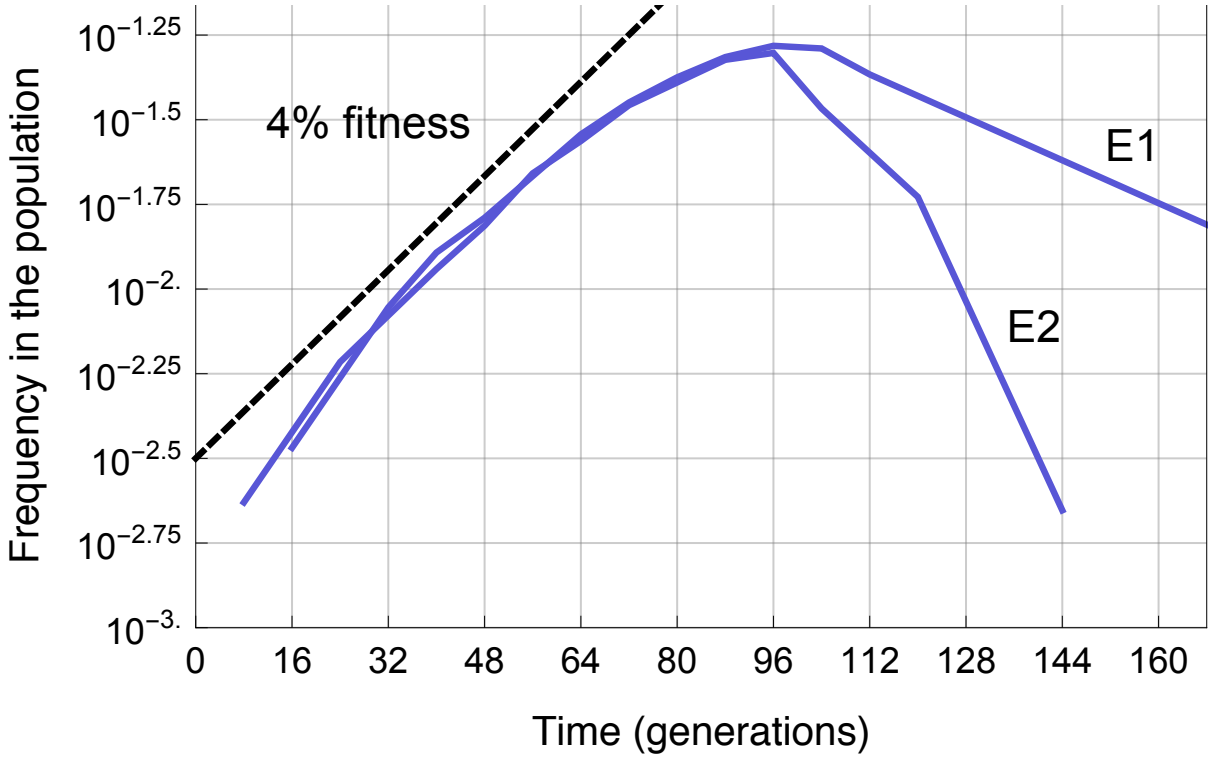
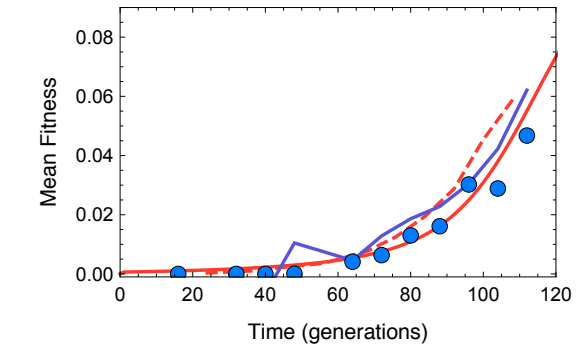
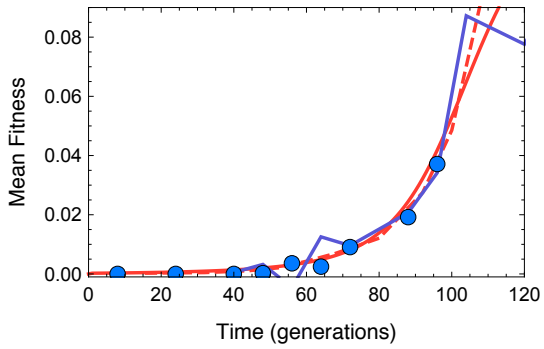


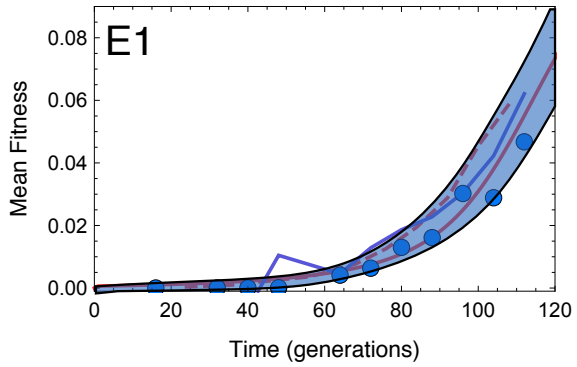
Figure 22: The aggregate trajectories of all pre-existing lineages whose fitness was inferred to be in the range $0.03 < s < 0.05$ (purple) tracks the expected exponential increase at a rate of $s = 0.04$ (dashed black line) until mean fitness begins to change. The deviation of the purple lines from the dashed black line gives an estimate of the mean fitness. Note the dramatic differences between the replicates at late times ($t > 96$): consistent with the mean fitness becoming stochastic at late times.

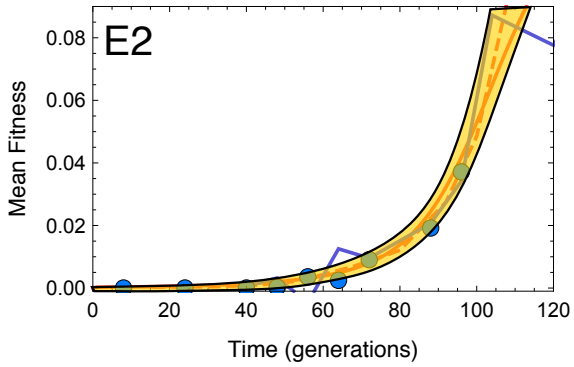


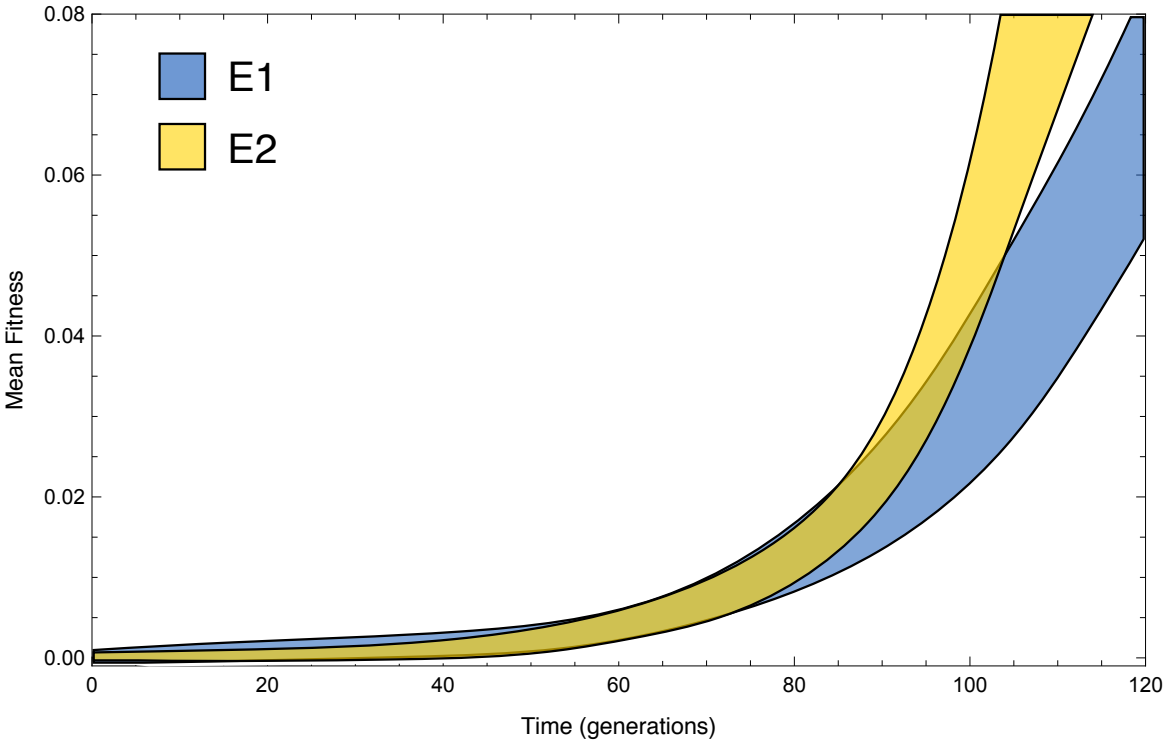
(a) E1



(b) E2







(c) Comparing the mean fitnesses between E1 and E2 with estimates for systematic errors.

Figure 23: **(a)** and **(b)**: plots comparing the mean fitness inferred from the deviation of the pre-existing class of $s = 0.04$ mutations from their exponential growth rate (purple lines). For comparison we show the mean fitness inferred from the decline of neutral lineages (blue circles), the mean fitness inferred from the expansion of adaptive lineages (red line) from *inferred* values of (s, τ) and the mean fitness from locally measuring the log-gradient of each expanding adaptive lineage (dashed red line). **(c)** Estimating the systematic errors on the mean fitness by fitting a region that encompasses the mean fitness trajectories from the 4 different methods (blue for E1 and yellow for E2), and, bottom, comparing these regions to one another.

6.5 Simulating the mean fitness from the inferred $\mu(s)$

To check that our inferred mean fitness and inferred $\mu(s)$ are consistent with one another, we performed simulations of the dynamics of the mean fitness using the inferred $\mu(s)$ distribution (see Figure 3B and 3E). We simulated these dynamics as follows:

1. We formed a set of putatively pre-existing lineages by selecting all lineage that were identified as adaptive in E1 and E2 and that had anomalously early establishment times ($\tau < -1/s$).

2. We simulated the contribution of these pre-exising lineages to the mean fitness accounting for errors in the establishment time inferences of order $1/s$ (we drew the establishment times for these pre-existing mutations from a normal distribution whose mean was the measured establishment time, with a standard deviation of $1/s$). These pre-existing mutations were assigned their measured fitness and then assumed to increase in frequency deterministically in the presence of a changing mean fitness (see below).

3. We then simulated the contribution of *de novo* mutations to the mean fitness by drawing mutations from the distribution of mutation rates across fitness effects that we inferred from E1 and E2:

$$f(s)\delta s = \frac{\exp\left(s(t-\tau) - \int_\tau^t \bar{x}(t)dt\right)}{s} \tag{53}$$

   where τ is a random variate from the distribution

$$\rho(\tau)d\tau = \frac{sd\tau}{\Gamma(N\mu(s)\delta s)} \exp\left(-sN\mu(s)\delta s\tau - e^{-s\tau}\right). \tag{54}$$

4. The mean fitness $\bar{x}$, is calculated at each generation and updated iteratively, having been initialized at $\bar{x}(t=0) = 0$.

The mean fitness trajectories predicted by these simulations are shown in Figure 24 as the gray trajectories. For comparison in red we show the mean fitness trajectory that results from the measured values of (s, τ) from the data in both E1 (lower of the two lines) and E2 (higher of the two lines). These are the same mean fitness trajectories as the solid lines in Figures 2A and 2C insets).

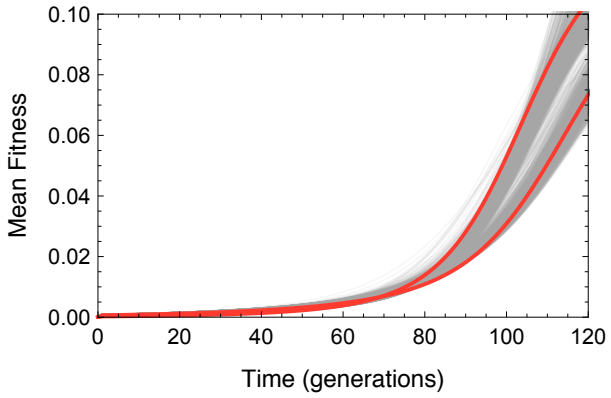


Figure 24: Simulated mean fitness trajectories (gray lines) compared to the measured mean fitness from each replicate (red lines) from 1000 simulations. Simulated trajectories were calculated using the inferred $\mu(s)$ from each replicate. We observe that both the measured E1 and E2 are similar to the simulated trajectories, which are themselves very similar to one another at early times, but show both stochastic differences (width of the gray clouds) and systematic differences (small separation between the gray clouds) at late times.

7 Inference of s and τ and their errors

For a given lineage trajectory (read number measurements over time) there are only two reasonable hypotheses, (called $\mathcal{N}$ and $\mathcal{A}$) that could explain the data:

$\mathcal{N}$ no adaptive mutation established in the lineage

$\mathcal{A}$ an adaptive mutation with fitness effect s occurred in the lineage and established to grow exponentially with an establishment time τ

(it is likely deleterious mutations do enter lineages, however since the deleterious mutation rate must be less than the bulk mutation rate, $U_d < U \sim 0.01$, only a small fraction of cells would accumulate deleterious mutations and these are selected against over a timescale of $1/s_d$, where s_d is the deleterious effect size, making such lineages indistinguishable from the null hypothesis $\mathcal{N}$). We want to compare the probabilities of these two hypotheses given the evidence contained in the read trajectory (data) i.e. $P(\mathcal{N}|\text{data})$ vs $P(\mathcal{A}|\text{data})$. To calculate the probability of each hypothesis given the data we will first calculate the likelihood of the trajectory given the hypothesis and then invert this using Bayes' theorem using prior probabilities for the two hypotheses.

7.1 Likelihood of neutral hypothesis, $\mathcal{N}$

When no mutation occurs, the expected change in reads between time points is determined entirely by competition against the increasing mean fitness $\bar{x}$. As outlined in Section 5 the expression for the probability of observing r_{t+T} reads at time $t+T$ conditioned on observing r_t and time t, $P(r_{t+T}|r_t, \text{no mutation})$, is well approximated by

$$P(r_{t+T}|r_t, \text{no mutation}) \approx \sqrt{\frac{(r_t e^{-T\bar{x}_t})^{1/2}}{4\pi\kappa_t r_{t+T}^{3/2}}} \exp\left[-\frac{\left(\sqrt{r_{t+T}} - \sqrt{r_t e^{-T\bar{x}_t}}\right)^2}{\kappa_t}\right] \tag{55}$$

where κ_t and $\bar{x}_t$ are the best fit values of the noise parameter and mean fitness between time point t and the subsequent time point $t+T$ as discussed in Section 6.1.
The inclusion of information from the first time point is slightly more subtle since for this time point we do not have a previous time point on which to condition. We do however have an accurate estimate of the frequency of the lineage at the zeroth time point f_0 as discussed in Section 5. We include this in the following way:

$$P(r_{t=0}|f_0, \text{no mutation}) \approx \sqrt{\frac{(R_0 f)^{1/2}}{4\pi\kappa_0 r_{t+T}^{3/2}}} \exp\left[-\frac{\left(\sqrt{r_{t+T}} - \sqrt{R_0 f}\right)^2}{\kappa_0}\right] \tag{56}$$

where κ_0 is set to 2 (typical of the other time points) and $\bar{x}_0 = 0$.

The full likelihood of the trajectory given the no-mutation hypothesis $\mathcal{H}1$ is then

$$\log\left(\text{Likelihood}(\text{trajectory}|\text{no mutation})\right) = \log P(r_{t=0}|f_0, s, \tau) + \sum_t \log P(r_{t+T}|r_t \;\&\; \text{no mutation}) \tag{57}$$

7.2 Likelihood of s, τ hypothesis, $\mathcal{A}$

If a mutation of fitness effect s and establishment time τ occurs, the growth of the number of beneficial-mutant cells is of the form

$$n(t) = \frac{c}{s - \bar{x}(\tau)} \exp\left(\int_\tau^t (s - \bar{x}(t'))dt'\right) \tag{58}$$

If the number of reads observed at the previous time point was r_t, and the read depth at this time point was R_t then approximately $(n(t)/N_e)R_t$ of the reads came from cells with a beneficial mutation and these will increase by a factor $e^{(s-\bar{x})T}$ at the next time point. The remanding, $r_t - (n(t)/N_e)R_t$, are derived from neutral cells and hence decrease by a factor of $e^{-\bar{x}T}$ at the next time point. Thus, the mean number of reads expected at the following time point is approximately:

$$\langle r_{t+T} \rangle \approx (n(t)/N_e)R_t e^{(s-\bar{x})T} + r_t - (n(t)/N_e)R_t \tag{59}$$

(note that if $(n(t)/N_e)R_t > r_t$, we set $(n(t)/N_e)R_t = r_t$ so that the entire lineage is predicted to contain beneficial cells). Therefore the probability of the data given the hypothesis is obtained from

$$P(r_{t+T}|r_t, s, \tau) \approx \sqrt{\frac{\langle r_{t+T} \rangle^{1/2}}{4\pi\kappa_t r_{t+T}^{3/2}}} \exp\left[-\frac{\left(\sqrt{r_{t+T}} - \sqrt{\langle r_{t+T} \rangle}\right)^2}{\kappa_t}\right] \tag{60}$$

where $\langle r_{t+T} \rangle$ is the expression in Eqn. 59. As in the case of the no mutation hypothesis we include the initial zeroth time point via

$$P(r_{t=0}|f_0, s, \tau) \approx \sqrt{\frac{(R_0 f)^{1/2}}{4\pi\kappa_0 r_{t+T}^{3/2}}} \exp\left[-\frac{\left(\sqrt{r_{t+T}} - \sqrt{R_0 f}\right)^2}{\kappa_0}\right] \tag{61}$$

The full likelihood of the data given the (s, τ) hypothesis is then

$$\ln\left(\text{Likelihood}(\text{trajectory}|s, \tau)\right) = \ln P(r_{t=0}|f_0, s, \tau) + \sum_t \ln P(r_{t+T}|r_t, s, \tau) \tag{62}$$

7.3 Prior for the neutral hypothesis, $\mathcal{N}$

The prior probability for the no-mutation hypothesis depends on U_b and how long the trajectory exists for. The analysis considered here lasts up to $t \sim 100$ generations by which time the probability of a lineage containing a beneficial mutation of *any* observable effect size is roughly

$$n_0 U_b \frac{s}{c} t \tag{63}$$

putting in $s \sim 0.05$, a typical' effect size, $2c \approx 3.5$ (see Section 5.4), $U_b(s > 2\%) \sim 10^{-5}$, $n_0 \sim 10^3$ and $t \sim 10^2$ we see that even by the end of the observation time the probability of a lineage having accumulated a beneficial mutation remains on the order of a few percent, and hence the prior probability of no mutation is close to unity i.e.

$$\text{Prior}(\text{no mutation}) \approx 1 \tag{64}$$

7.4 Prior for the (s, τ) hypothesis, $\mathcal{A}$

The prior probability of a mutation occurring in the interval ds around s is largely unknown since it is the very distribution we are attempting to infer from the data. However, previous attempts to determine this distribution have yielded some prior knowledge which informs our prior:

- There are generally fewer mutations of larger effect, hence $\mu(s)$ should be a decreasing function. [15]
- Total beneficial mutation rates, U_b in yeast are in the range $U_b \in \left[10^{-6} - 10^{-4}\right]$ [16, 17]
- Selection coefficients in yeast evolution experiments are typically in the percent range [16]

Given these considerations we elected to use a prior in s of the form

$$\mu(s)ds = U_b \frac{ds}{\bar{s}} \exp\left(-s/\bar{s}\right) \tag{65}$$

where

$$\int_0^\infty \mu(s)ds = U_b = 10^{-5} \tag{66}$$

with $\bar{s} = 0.1$. This prior is intentionally broad, reflecting our lack of knowledge about the range of fitness effects we expect to see in the system. However the exact form of the prior at large fitness is not very important as inferences for mutations with such large fitness effects are highly constrained by the data.

This prior distribution over s can — in principle — be used to determine the prior probability of the mutation establishing in the interval $d\tau$ around τ. That is, $\mu(s)$ sets $\rho(\tau)$. However to do this assumes that the process of establishment of single beneficial mutations results from a constant feeding process from the pool of $\sim n_0$ neutral cells as outlined in Sections 7. We chose not to use the distribution of establishment times as outlined in Eqn. 7 as a prior on τ because it does not account for the possibility of mutations arising in the period of common growth prior to the beginning of the growth-bottleneck cycles that were discussed in Section 10. Since we have less concrete information about the environment, mutation rates and effect sizes of mutations in this period of prior growth, this lack of information should be reflected in our prior. We therefore chose a uniform prior over τ of the form

$$\rho(\tau)d\tau = \frac{d\tau}{\Delta\tau} \qquad \text{for } -150 < \tau < 100 \tag{67}$$

where $\Delta\tau = 250$. We intentionally included the possibility of very negative establishment times because, as discussed in Section 10 we expect a large number of pre-existing mutations to have substantially negative establishment times.

7.5 Bayesian Posterior

For each trajectory we calculate the ratio of the the posterior probabilities for $\mathcal{H}1$ and $\mathcal{H}2$:

$$r(s,\tau) = \frac{P(\mathcal{H}2|\text{data})}{P(\mathcal{H}1|\text{data})} = \frac{\text{Prior}(s,\tau) \times \text{Likelihood}(s,\tau)ds\ d\tau}{\text{Prior(no mutation)} \times \text{Likelihood(no mutation)}} \tag{68}$$

over ranges

$$0 \leq s \leq 0.4 \tag{69}$$

$$-150 \leq \tau \leq 100 \tag{70}$$

using bin widths of $\delta s = 0.005$ and $\delta\tau = 1$.

If, for any range of (s,τ) this ratio the hypothesis of the particular (s,τ) in invertval $(\delta s, \delta\tau)$ is more likely than the neutral hypothesis ($r > 1$) we classify the lineage as adaptive. If the lineage is classified as adaptive, the position in (s^*,τ^*) space at which r is maximized is used as our best estimate for fitness effect and establishment time of the mutation.

Figure 25 plots some of these posterior probabilities. Each row of the plot is a particular barcode. The columns are labelled in the figure. Note here the labels 2M3 = E1 and 4M3 = E2. The plots of the posterior probabilities for the barcode in both replicates E1 and E2 are shown in the final two columns. The plots of the data are in the 11th column and show the trajectory data from replicate E1 (black points) and E2 (white points) compared to the the predicted trajectory of the most probable hypothesis (s^*,τ^*) (solid black for E1, dashed black line for E2) and the predicted trajectory of the the no-mutation hypothesis (solid gray line E1, dashed gray line E2).

Despite having to chose a prior for the purposes of the inference, our results are largely insensitive to the exact form of the prior we use: for most adaptive lineages the data (contained in the likelihood) constrains the range of possible (s, τ) to a very narrow region. Mutations for whom evidence is weak (i.e. their posterior probability ratio r is close to unity) are inevitably affected by the prior expectation. However, these are mostly lineages whose size does not increase much and hence have little effect on the population dynamics or our quantitative results.

One could use the inferred $\mu(s)$ (see Section 11) as a prior and repeat the process of inference to check for self consistency, iterating a number of times if needed. However because the prior does not grossly affect most of our inferences the gains from performing this iteration are marginal. Another consideration is that as our inferences become more confident the uncertainties in s, τ are more strongly affected by systematic differences in mean fitness and in the inherent uncertainties between the establishment time $\tau = \tau_{est}$ and the true mutation occurrence time τ_{mut} than they are by uncertainties caused by the prior.

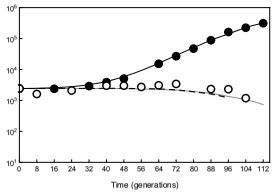
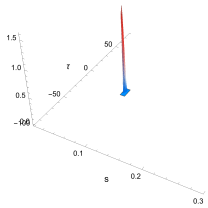
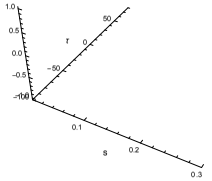
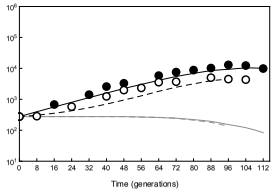
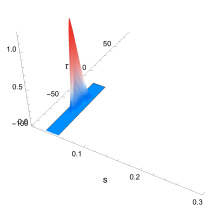
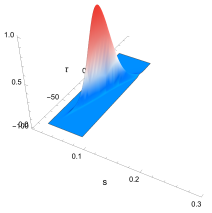
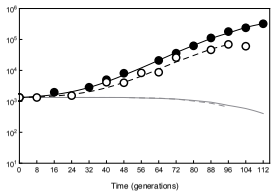
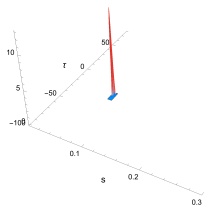
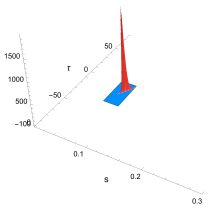
BC	Log(Pr[b]/Pr[n])_2M3	Log(Pr[b]/Pr[n])_4M3	s_2M3	δs_2M3	s_4M3	δs_4M3	τ_2M3	δτ_2M3	τ_4M3	δτ_4M3	Trajectory Plots	Posterior Plot_2M3	Posterior Plot_4M3
22011	5864.23	-9.00841	0.088832	0.000879472	0.00001	0.01	-16.7756	1.15536	0	5			
22469	82.4711	2.89351	0.0417933	0.00235543	0.0520239	0.00657389	-60.8967	8.40216	-22.1391	13.2153			
23688	5698.06	423.347	0.0812479	0.000758688	0.0812822	0.00290723	-27.9064	1.18145	-18.9423	3.66083			

Figure 25: A table of the data resulting from the analysis of each barcode trajectory. The columns are (i) barcode ID (ii) the log ratio of the most likely $\mathcal{A}$ hypothesis to the neutral hypothesis $\mathcal{N}$, i.e. $\ln(r(s^*, \tau^*))$ in replicate E1. A positive value indicates that the maximum posterior in a range $\delta s = 0.005$ and $\delta\tau = 1$ around (s^*, τ^*) is more probable than the null hypothesis of no-mutation. (iii) $\ln(r(s^*, \tau^*))$ for replicate E2. (iv) The most probably estimate of the fitness effect of the mutation s^* in replicate E1, (v) The error, σ_s, in the fitness estimate in E1, (vi) and (vii) the corresponding s^* and σ_s for replicate E2. (viii) The most probable estimate for the establishment time τ^* in E1 (ix) the error in the estimate of establishment time in E1 σ_τ. (x) and (xi) the corresponding τ^* and σ_τ for E2. Column (xii) is a plot of the trajectory data from E1 (black points) and E2 (white points) compared to the the predicted trajectory based on the hypothesis (s^*, τ^*) (solid black for E1, dashed black line for E2) and the predicted trajectory of the the no-mutation hypothesis (solid gray line E1, dashed gray line E2). The final columns show plots of the posterior probability ratio $r(s, \tau)$ normalized so that the maximum posterior $r(s^*, \tau^*) \approx 1$. The range of each plot is $\pm 5\sigma_s$ along the s axis and $\pm 5\sigma_\tau$ along the τ axis. Absence of a posterior probability surface indicates that the lineage was not identified as adaptive.

7.6 Visualizing barcode trajectories by fitness

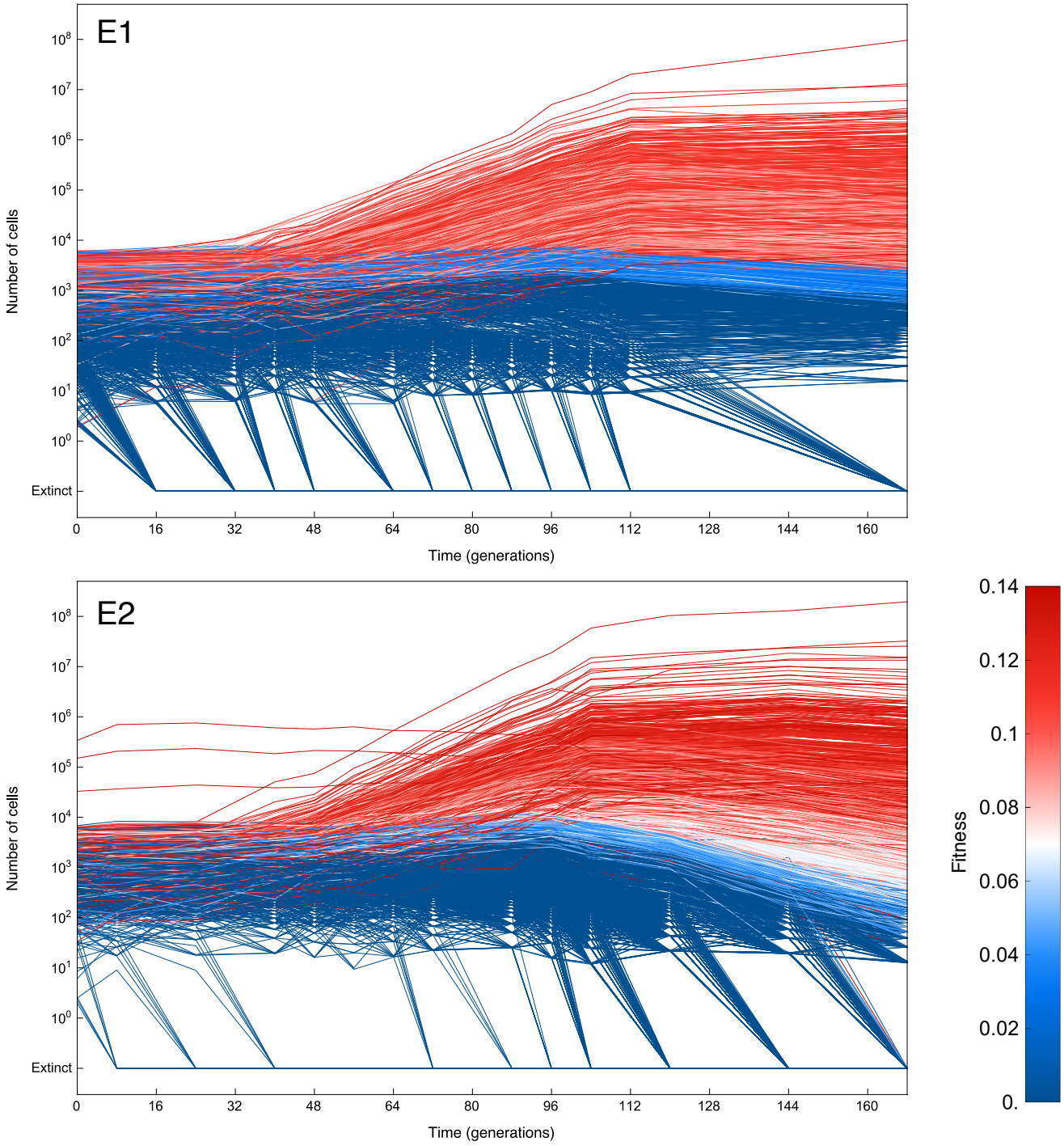


Figure 26: Sampled trajectories from both replicates colored according to the inferred fitness of the mutation in each barcode.

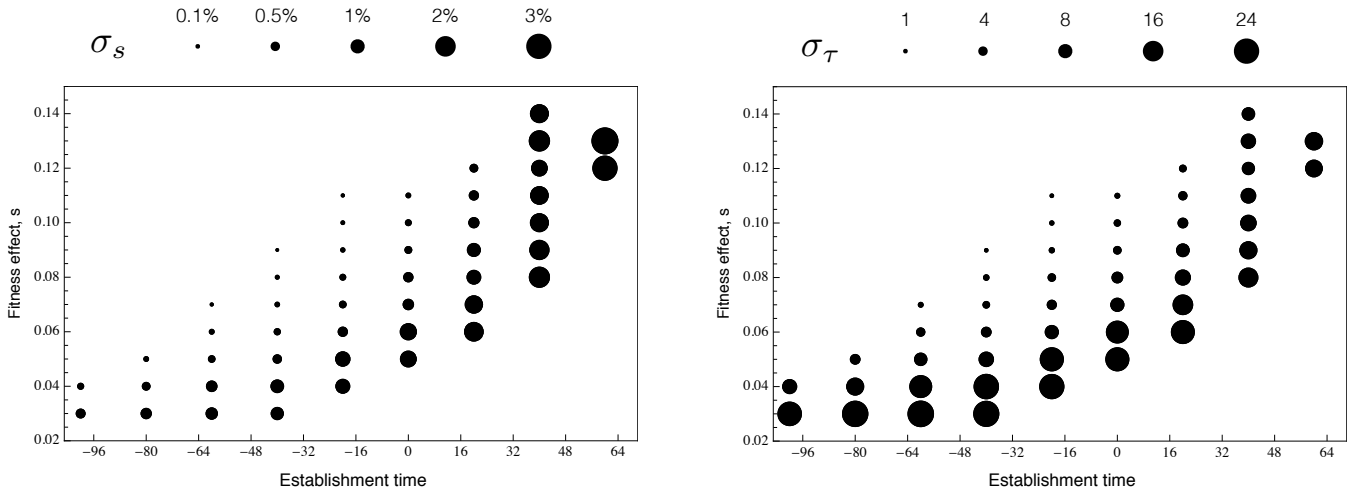


Figure 27: Plot of the mean errors in s (left plot) and τ (right plot) binned by different (s, τ) with bin widths of $s = 0.01$ and $\tau = 20$). The error are depicted by the size of the black circles (scale above each plot). The errors depend on when the mutation arose and how large its fitness effect is, however errors in s are typically $\sim 0.5 - 1\%$ while errors in τ are typically $5 - 15$ generations

7.7 Errors in s and τ

Errors on the best estimate (maximum posterior) fitness effect and establishment time (σ_s, σ_τ) follow straightforwardly from $r(s, \tau)$. Intuitively the errors in both are related to how rapidly the posterior probability decays away from (s^*, τ^*) and thus to the breadth and thickness of the peak. We calculate these errors as follows

- We calculate the curvature matrix around the maximum probability

$$K = \left(\begin{array}{cc} \partial_s^2 \ln r & \partial_\tau \partial_s \ln r \\ \partial_s \partial_\tau \ln r & \partial_\tau^2 \ln r \end{array} \right) \Bigg|_{(s^*, \tau^*)} \tag{71}$$

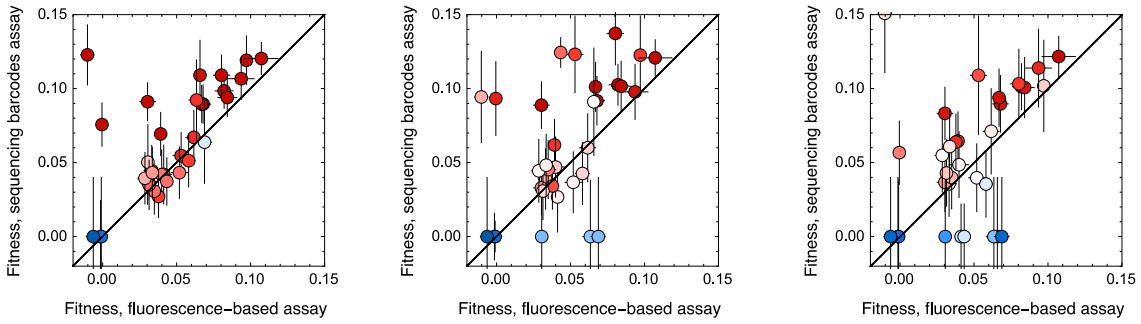
- Find the two eigenvectors $(\mathbf{v_1}, \mathbf{v_2})$ and corresponding eigenvalues λ_1, λ_2 of this matrix. The eigenvectors are the directions of principle curvature and their eigenvalues the respective principle curvatures in those directions. Moving along the direction $\mathbf{v_1}$ away from the maximum posterior probability at (s^*, τ^*), the posterior probability decays like a Gaussian with standard deviation $1/\sqrt{\lambda_1}$ and similarly for $\mathbf{v_2}$ with corresponding standard deviation $1/\sqrt{\lambda_2}$.

- The magnitude of $(\mathbf{v_1} \cdot \hat{\mathbf{s}})/\sqrt{\lambda_1}$ and $(\mathbf{v_2} \cdot \hat{\mathbf{s}})/\sqrt{\lambda_2}$ (where $\hat{\mathbf{s}}$ is the unit vector $(1, 0)$) give the errors in fitness effect, δs along the two principle directions. We define the error in fitness as the maximum of these two values. Similarly $(\mathbf{v_1} \cdot \hat{\tau})/\sqrt{\lambda_1}$ and $(\mathbf{v_2} \cdot \hat{\tau})/\sqrt{\lambda_2}$ give the errors in establishment time $\delta\tau$ along the two principle directions, the maximum of which is taken as the error.

The errors in s and τ depend on the specific trajectory in question, and, more generally on (s, τ). Figure 27 shows how the errors σ_s and σ_τ depend on (s, τ). We note that the typical errors in the inference of $\tau \approx 15$ generations are smaller than the inherent uncertainties between the establishment time $\tau = \tau_{est}$ and the true mutation occurrence time τ_{mut} that can differ by a few multiples of $\sim \pm 1/s$ (see Figure 7)

8 Systematic Errors

In addition to errors associated with the width of the posterior distribution, there are also factors that likely contribute systematic errors to estimates of (s, τ). Here we discuss the various factors that contribute to systematic errors and estimate their magnitudes.



(a) Times up to 88 **(b)** Times up to 72 **(c)** Times up to 64

Figure 28: Scatter plots of the fitnesses of 33 clones that were picked at generation $t = 88$ from replicate E2 and fitness assayed in a fluorescence competition experiment as described in Section 2. The inferred fitnesses agree very well for clone with low fitness, however there is a systematic bias for clones at high fitness. **(a)** shows plots inferred values from sequencing up to $t \leq 88$. The systematic bias is not reduced if the analysis is restricted to earlier times, e.g. $t \leq 72$ in **(b)** or $t \leq 72$ in **(c)**. This suggests that the bias cannot be fully explained by systematic errors in mean fitness inferences at late times.

Mean Fitness errors. The fitnesses quoted in this work are relative to the fitness of the ancestral strain. As discussed in Section 7, this requires accurately accounting for the mean fitness of the population. Mean fitness measurements naturally have errors associated with them and, as can be seen from Figure 23, these errors are $\approx 1\%$ at times $t > 64$ and $< 1\%$ for $t < 64$.

Uncertainty in the inferred mean fitness at later times is expected for a number of reasons: (i) there are fewer neutral lineages making it more challenging to track the decline of them (see Section 6.1). At later times these neutral lineages are also more likely to contain beneficial mutations and therefore to potentially underestimate the true mean fitness. (ii) The diversity of barcodes is reduced at later times and expanding beneficial lineages become large enough to potentially contain multiple beneficial mutations. This worsening frequency resolution at late times makes it more difficult to accurately capture the fitness of each expanding lineage. Unfortunately, these late times are also when many of the beneficial mutant lineages become large enough to enable accurate inferences of s: thus for these, the errors from the uncertainties in the mean fitness will be largest.

To quantify how these systematic errors in mean fitness may affect our fitness estimates, we consider two approaches. First, we compare the fitness inferred via barcode sequencing to the fitnesses inferred from a fluorescence based competition assay (see Section 2 for details on how the fluorescence assay was performed). Second, we use "pre-existing" lineages — lineages likely containing beneficial mutations that arose prior to the separation of the two replicates that subsequently established in both replicates— as useful checks on our inferences since we know it is likely that in most cases these lineages contain mutations of the same identity, and hence of the same fitness.

8.1 Comparison with fluorescence assay fitness.

Figure 28 shows scatter plots of the fitnesses of 33 clones (from replicate E2 only) as measured via the fluorescence based assay (x-axis) and the barcode sequencing inferences (y-axis) using data up to 88 generations (a) up to 72 generations (b), and, up to 64 generations (c). While there is very good agreement for the clones whose fitnesses are in the $0.03 < s < 0.05$ range, the higher fitness clones have a small systematic bias with fitnesses from barcode sequencing being systematically higher than the fitness measured via fluorescence. One potential explanation for this bias is that at later times, a small systematic error in the inferred mean fitness of the population could cause a systematic shift of fitnesses inferred from barcode sequencing. However, restricting the analysis to earlier times —when systematic errors in mean fitness are $< 1\%$ — does not significantly affect the observed bias. (It does however make the inferences more noisy, because they are based on less data).

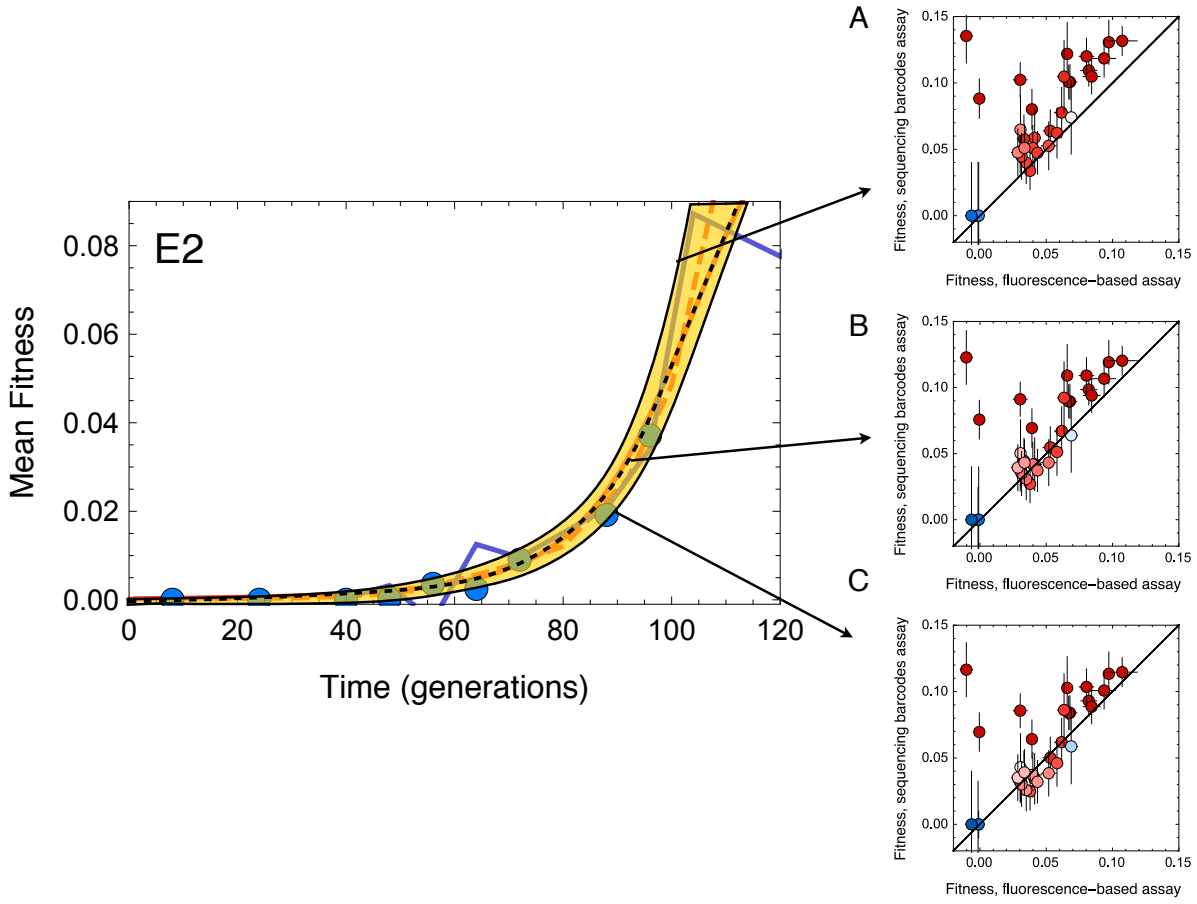


Figure 29: Using three different mean fitness trajectories, that approximate the magnitude of systematic errors (see Section 6.4), we re-inferred the fitnesses of the 33 clones that were fluorescently assayed from replicate E2. Here the analysis is again restricted to $t \leq 88$. Error bars on each fitness measurement is shown via black lines behind each of the points. While there can be some systematic effect of the mean fitness on the observed bias it seems unlikely that the bias in the fitter clones is fully explained away via errors in mean fitness.

To look further into the role of the mean fitness on our inferences we used an upper and lower bound for the mean fitness from replicate E2 by estimating the magnitude of systematic errors in the mean fitness (Figure 29, yellow region, see 6.4) and re-inferred the fitnesses of the 33 assayed clones for high (A) medium (B) and low (C) mean fitness estimates. While a lower mean fitness does reduce the bias for the higher fitness clones, it lowers the estimates for the low fitness clones by too much. Given that errors in the mean fitness for early times $t < 88$ are small ($<1\%$) it seems implausible that errors in the mean fitness can fully explain the observed bias.

To further check this, we examined plots of the measured barcode abundance data (Figures 30-33 data points) for each of the 33 picked clones compared to the predicted trajectories based on our inferences (Figures 30-33 solid yellow line, the yellow dashed line is the neutral trajectory). As a guide, we also plot the gradient of the trajectory that would be expected from the fitness from the fluorescent assay (Figures 30-33 black dashed line). In all cases the predicted trajectories track the measured ones very closely. In the majority of cases the fitness measured from the fluorescent assay also has the correct gradient. However there are clear cases where the fitness measured from the fluorescent assay could not possibly explain the observed trajectory. Barcodes #8825 and #29375 are clearly beneficial (and substantially so) from the sequencing trajectories but were measured to be neutral via the fluorescent assay. These discrepancies are

very likely due to the fact that the picked clones was one of the remaining neural cells in the expanding lineage. Indeed, given that the expanding mutations in most lineages are only ~ 10 fold more frequent than the remaining neutral cells, we should expect some of these from a sample of 33 clones.

Other discrepancies between barcode-sequencing fitness and the fluorescent fitness however are more of a mystery. Barcode #14280 is an example where there is a clear discrepancy between the two fitnesses that could not be explained by small systematic differences in mean fitness (barcodes #25531 and #63215 are less extreme examples). These could potentially be multiple mutants causing the lineage to expand by more than can be explained by the first mutation or signs of ecological effects coming in.

From these considerations we conclude that while a small systematic error in mean fitness could account for part of the observed bias, it cannot account for all of it. The emergence of multiple mutants in some of the lineages could explain some of the discrepancies, it is unlikely they are the whole story because the bias exists at early times, when multiple mutants are expected to be very rare. Another possibility is that there could be genuine differences in the growth rate of the same mutation in the different assays. This is plausible given that we know the growth rates in the fluorescent assay are very sensitive to small differences in conditions. There are indeed non-trivial differences in conditions between the two assays: for example there are large differences in frequencies at which each of the clones is present (in the fluorescence-based assay the clone is present at $\sim 10\%$ while in the sequencing assay even the high fitness are typically at the 0.01% scale) and differences in the presence /absence of other expanding beneficial mutant subpopulations.

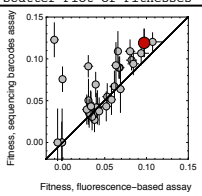
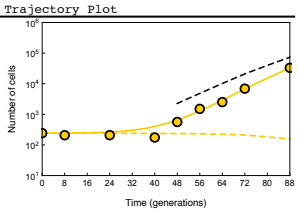
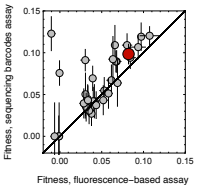
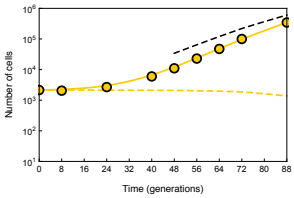
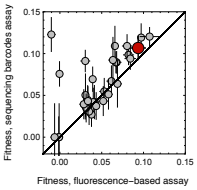
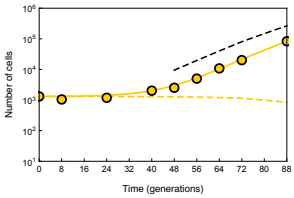
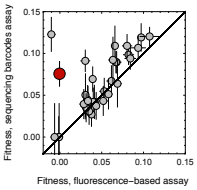
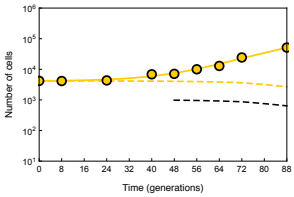
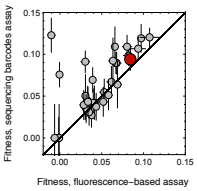
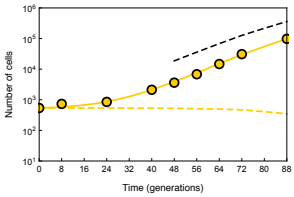
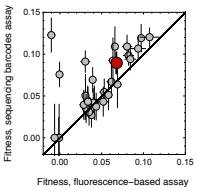
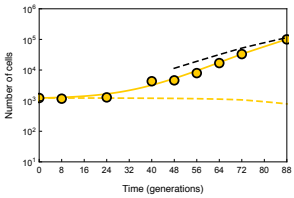
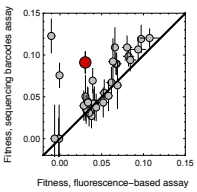
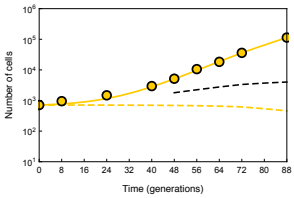
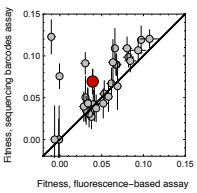
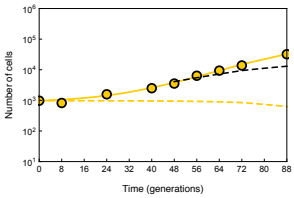
BCID	Fluoro s	Seq s	log[P(b)/P(n)]	Scatter Plot of Fitnesses	Trajectory Plot
415	0.0971749	0.119144	329.044		
2468	0.0820773	0.0984566	3007.09		
7961	0.0936212	0.106694	766.842		
8825	-0.000462491	0.0756987	273.412		
9888	0.0842182	0.0940048	808.966		
10 851	0.0678635	0.0895449	797.939		
14 280	0.030357	0.0911299	913.809		
18 501	0.0391938	0.0693557	189.941		

Figure 30: Fluorescence and barcode sequencing fitness for the 33 picked clones (each highlighted). Yellow data points show the measured abundance in E2, the yellow line is our inferred trajectory, the dashed yellow line the expected neutral trajectory. The black dashed line is the gradient that would be predicted by the fluorescent fitness.

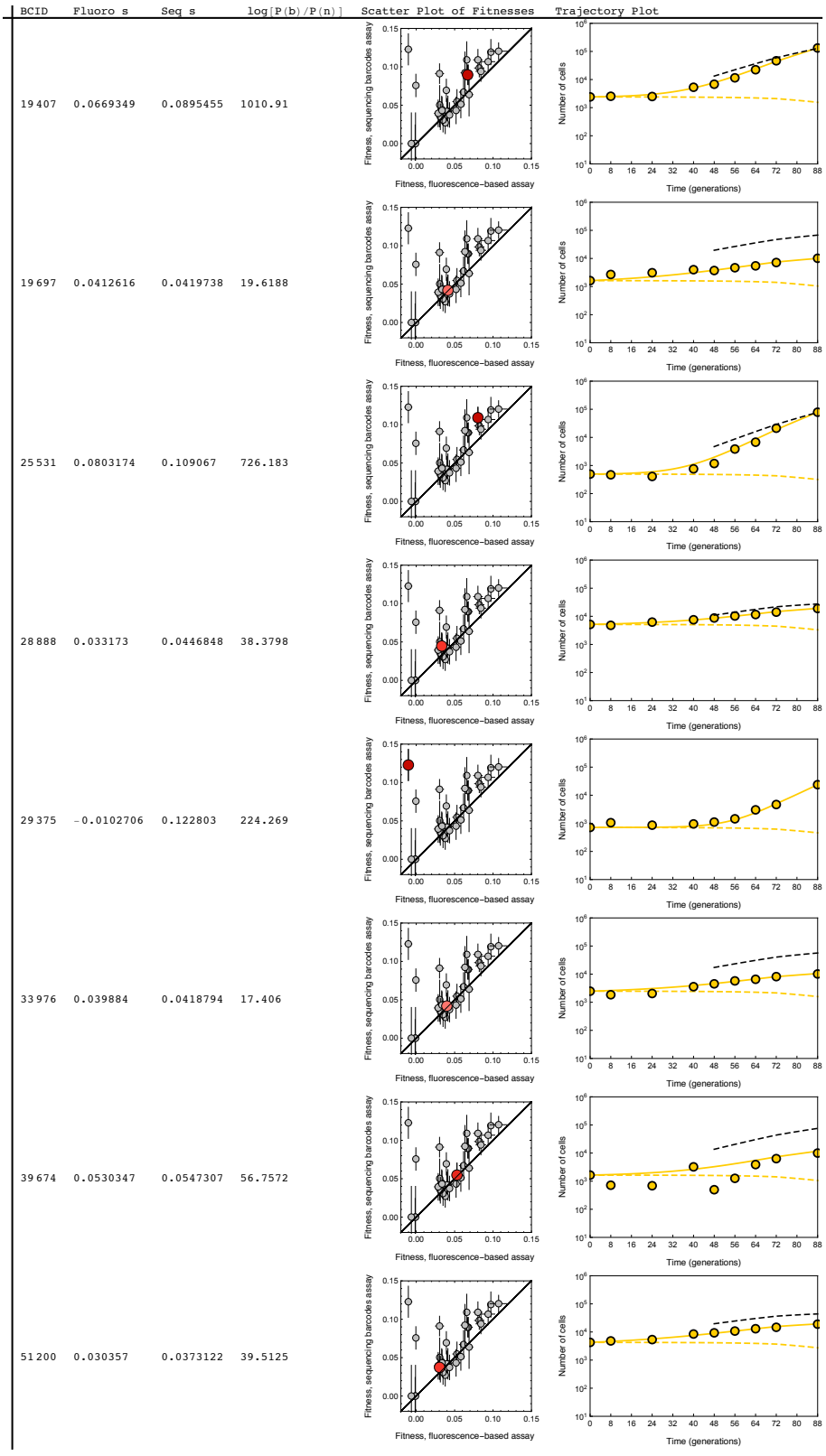


Figure 31: Fluorescence and barcode sequencing fitness for the 33 picked clones (each highlighted). Yellow data points show the measured abundance in E2, the yellow line is our inferred trajectory, the dashed yellow line the expected neutral trajectory. The black dashed line is the gradient that would be predicted by the fluorescent fitness.

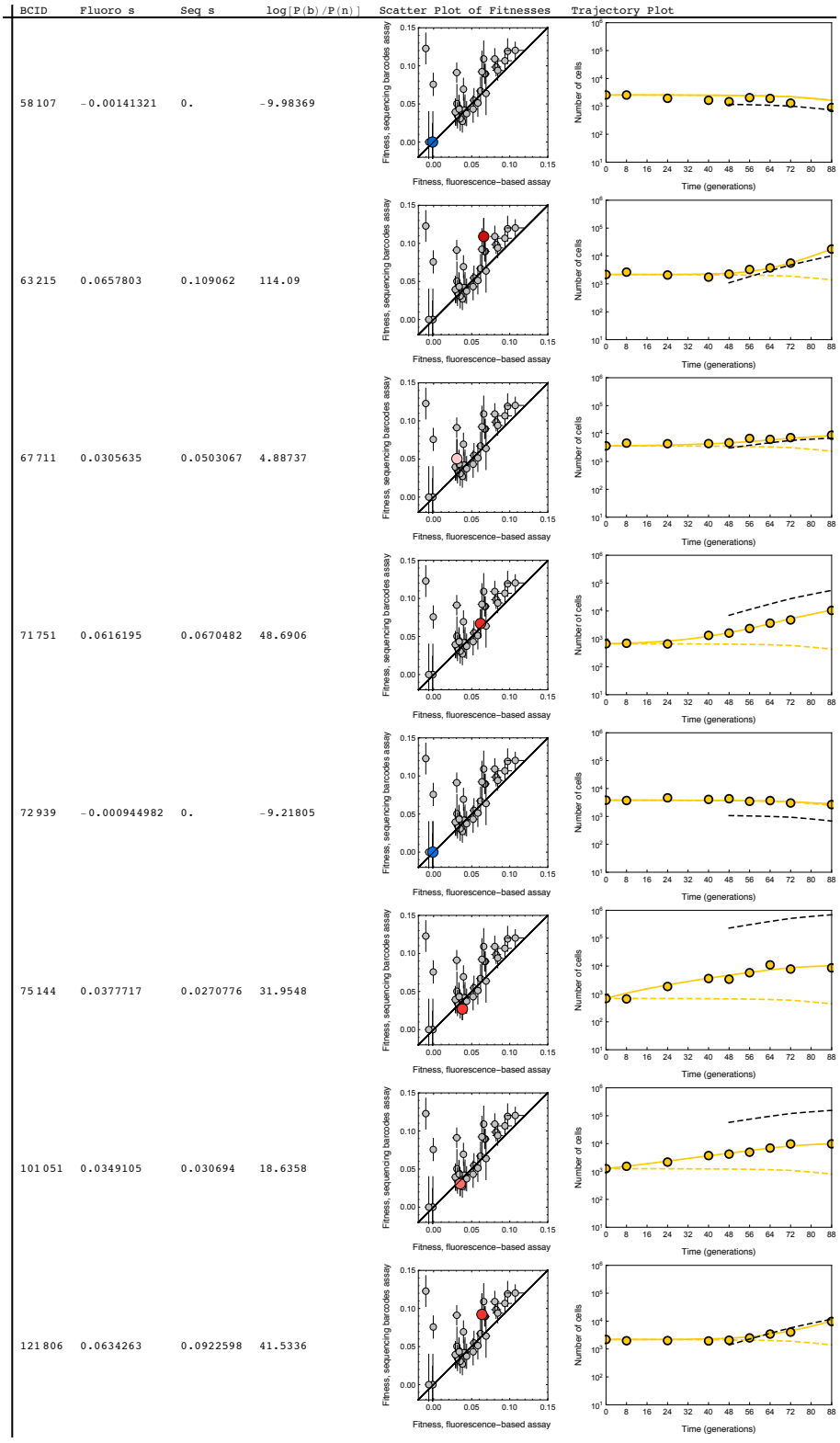


Figure 32: Fluorescence and barcode sequencing fitness for the 33 picked clones (each highlighted). Yellow data points show the measured abundance in E2, the yellow line is our inferred trajectory, the dashed yellow line the expected neutral trajectory. The black dashed line is the gradient that would be predicted by the fluorescent fitness.

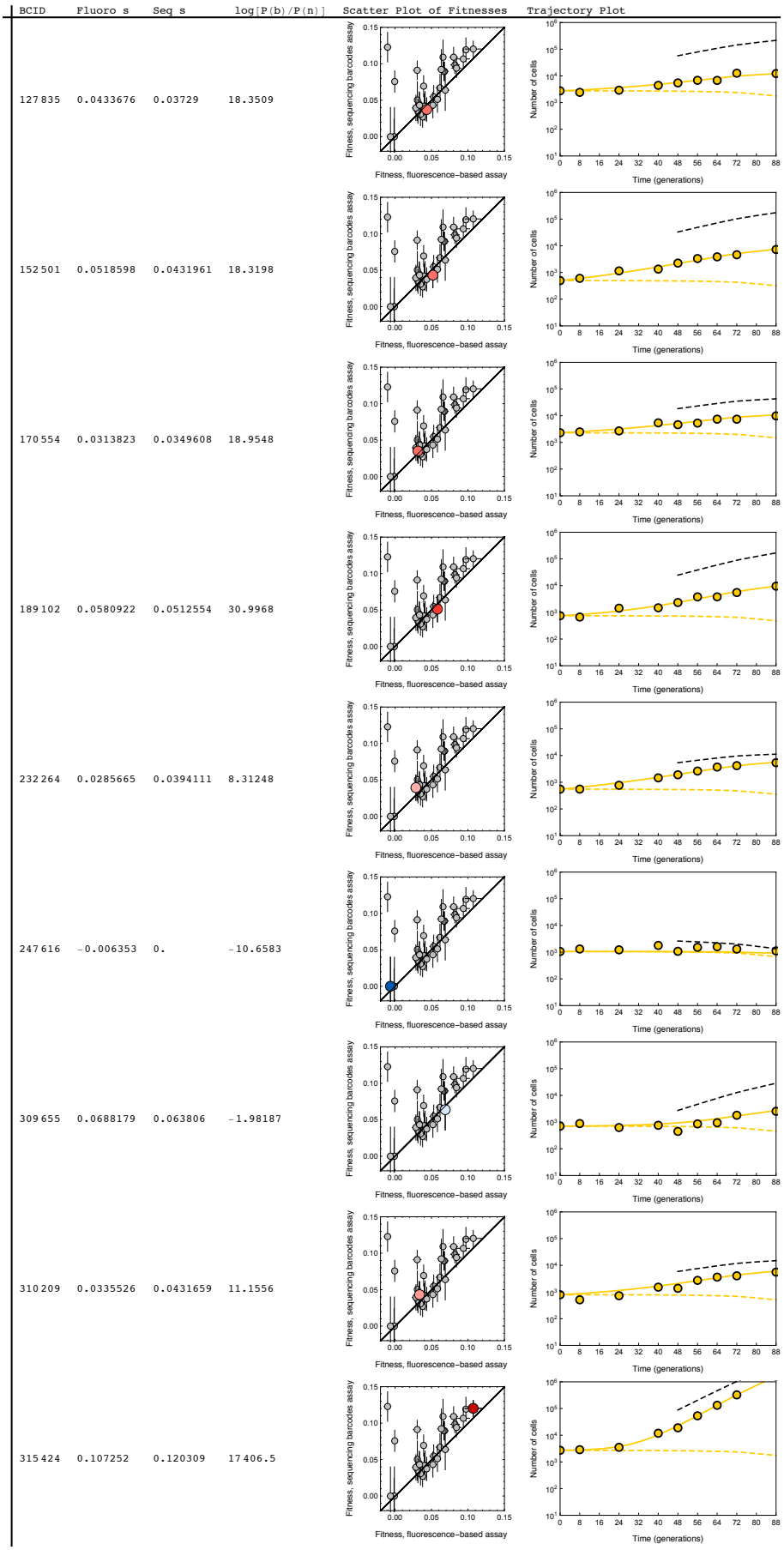


Figure 33: Fluorescence and barcode sequencing fitness for the 33 picked clones (each highlighted). Yellow data points show the measured abundance in E2, the yellow line is our inferred trajectory, the dashed yellow line the expected neutral trajectory. The black dashed line is the gradient that would be predicted by the fluorescent fitness.

8.2 Using Pre-existing mutations to verify fitness and establishment times

Pre-existing mutations (see Section 10) that arose prior to the splitting of the two replicates offer a very useful check on systematic errors. Since they contain the same beneficial mutation, these pre-existing lineages should have the same inferred fitness, and a similar (though not necessarily the same) inferred establishment time. To check this we plotted the fitnesses (Figure 34A) and establishment times (Figure 34B) of all lineages that were identified as adaptive across both replicates. As expected most fitness measurements and establishment times broadly agree. For mutations in this $0.03 < s < 0.05$ range the systematic differences between E1 and E2 are very small (< 0.005 absolute fitness difference).

However, as we saw in the previous discussion on fitnesses measured by a fluorescence assay, the majority of the bias appears to come from high-fitness clones. Therefore we also consider the rarer lineages that acquire large effect mutations in the range $0.06 < s < 0.15$, that are also likely pre-existing. To isolate a set of putatively pre-existing large-fitness-effect lineage we selected all lineages that:

- Are identified as adaptive in both E1 and E2
- Have fitness effects $s > 0.05$
- Have negative establishment times $\tau < 0$

A scatter plot of the fitnesses of these lineages inferred from E1 and E2 (Figure 35) shows a good correlation at late times ($t = 88$), though with a slight systematic offset (red line versus grey line) of magnitude $\delta s \approx 0.008$. There is little evidence for this systematic offset at early times (see panels with $t = 48$ and $t = 64$). This suggests therefore that the systematic offset is *not* likely to be an artifact. Rather it suggests that larger fitness effect mutations do have a slightly larger fitness advantage over neutral cells in the replicate E2 compared to E1. One possible explanation for this is ecological effects could be becoming significant at this time. This is plausible given that around this time ($t \sim 96$) mutant cells becomes the majority type of cell in the population. Another possibility is there is a very slight difference in the experimental conditions: from the fluorescent assay fitness measurements, we know that subtle differences — which we had to be very careful to avoid — can lead to far larger changes in fitnesses.

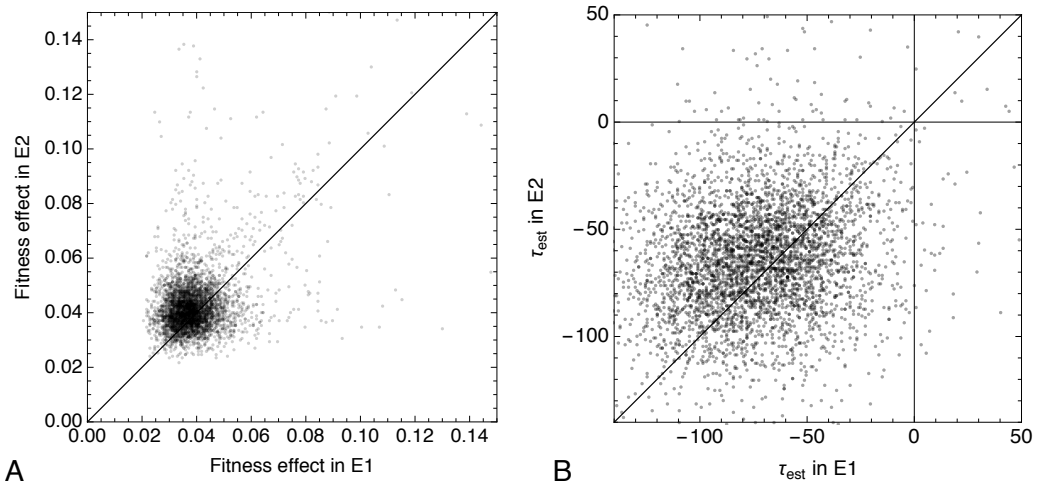


Figure 34: (A) The inferred fitnesses for all lineages that were identified as adaptive across both replicates E1 and E2. There is a very small systematic error for low fitness clones ($s \approx 4\%$) of less than 0.5%. Higher fitness clones seem to show larger deviations. (B) The inferred establishment times for all lineages that were identified as adaptive across both replicates E1 and E2. Note that the correlations is expected to be worse for establishment times given the inherent stochasticity during growth being independent in each replicate.

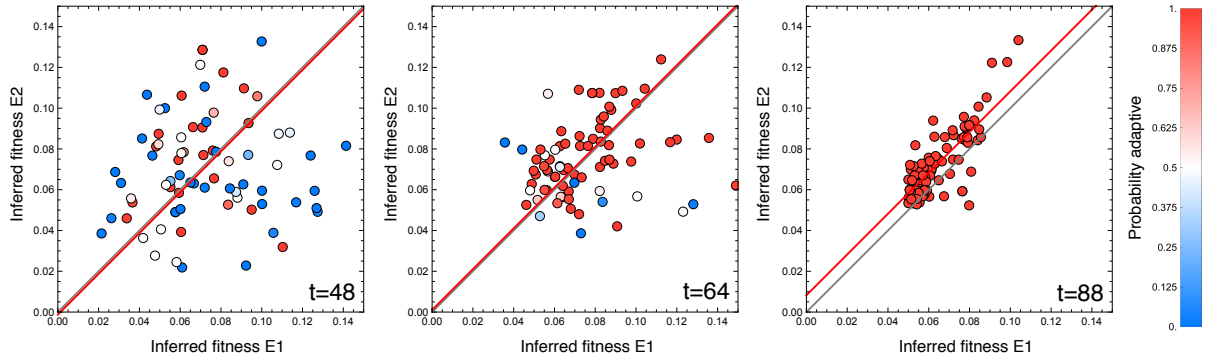


Figure 35: Scatter plots of the inferred fitnesses in replicates E1 and E2 for all lineages that contain putatively high fitness pre-existing mutations ($s > 0.05$, $\tau < 0$, and identified as adaptive across both replicates). Each panel shows the correlation in fitness that would result if only data up to the denoted time point were used. The color indicates the probability that the barcode is adaptive by that time (averaged across both replicates). The gray line is $x = y$, while the red line is $y = x + \epsilon$ with ϵ the least square fit to the data to determine if there is evidence for a systematic offset. At early times ($t = 48$ and $t = 64$), while the correlation is worse (as would be expected), there is no evidence of a systematic difference in fitness between the two replicates. At later times however ($t > 88$) there is some evidence that large fitness-effect mutations have a larger advantage in E2 compared to E1. This is corroborated in the correlation plot in Figure 2D of the main paper.

9 Detectability limits and small effect mutations

Consider a mutation with fitness effect s that enters the population at time t_m and — in the absence of any mean fitness increase — is destined to grow exponentially. The probability of this mutation establishing in the population and being detected as beneficial depends on how long it takes for the mutation to establish and how quickly the mean population fitness increases. If, by the time the mutation establishes in the population, the mean fitness has increased to above s, then the mutation begins to be outcompeted before it has been able to grow exponentially. As we outline in Section 9.1 this prohibits weak effect mutations from ever having much of an impact on the population dynamics and in effect places a lower bound on the fitness effects that can be detected in an adapting population. This limit emerges from the changing mean fitness rather than from detectability limits of the lineage tracking method.

In addition to limits on detectability of small fitness effects imposed by the adapting population, there is also a limit that emerges because of initial lineage sizes of n_0 cells. Since we measure the abundance of the entire lineage, not that of the individual mutation, a beneficial mutation must reach of size of $\sim n_0$ cells before its effect can be easily detected. This limit is discussed in Section 9.2.

9.1 Limits on s imposed by clonal interference

A mutation that occurs at time $t = \tau_{mut}$, that is destined to establish, and that competes against a mean fitness of $\bar{x} = 0$ grows as

$$n(t) = \frac{c\nu}{s}\left(e^{s(t-\tau_{mut})} - 1\right) \qquad \text{with} \qquad \nu \sim 1 \tag{72}$$

Where ν is a random variate from an exponential distribution (see Section 2), s the fitness effect and c the half the variance in offspring number. Soon after occurring, for $t - \tau_{mut} < 1/s$, the *average* growth of a mutation that has not gone extinct is roughly linear in time

$$\langle n(t) \rangle \sim c(t - \tau_{mut}) \tag{73}$$

and so the time taken to reach the establishment size of $\sim c/s$ is approximately $\sim 1/s$ generations, with variations around this of the same order, with a tail to late establishments. For weak effect mutations this

delay in reaching establishment size can thus be substantial and long enough for the mean population fitness to increase appreciably.

The mean fitness plots from experiment E1 and E2 in Figures 20a and 20b show that in both experiments it takes $\sim 80-100$ generations for the mean fitness to increase by a few percent. Therefore, even if a mutation with fitness effect $s = 0.01$ occurs immediately ($\tau_{mut} = 0$), it is unlikely to establish in the population before being outcompeted by the increasing mean fitness and therefore such mutations are unlikely to ever truly establish. More generally, for a mutation of effect size s to establish it would have to occur at least $\sim 1/s$ generations before the mean fitness $\bar{x}$ surpasses s. In terms of its occurrence time τ_{mut} this means

$$\tau_{mut} < t(\bar{x} = s) - 1/s. \tag{74}$$

This sets a time limit on detectability that arises from the inherent clonal interference. To establish in the populate a mutation must occur early enough so that it has time to grow exponentially before the mean fitness outcompetes it. This can also be cast in terms of a limit on fitness effect, s given an occurrence time

$$s \gtrsim \frac{1}{t(\bar{x} = s) - \tau_{mut}} \tag{75}$$

where $t(\bar{x} = s)$ is the time at which the mean fitness $\bar{x} = s$.

9.2 Limits on s imposed by the initial lineage size

The limit on the detectability of mutations in the previous section 9.1 assumes that in order to be detected a mutation only has to establish i.e. reach a size of $n \gtrsim 1/s$. However in the lineage tracking experiment, lineages also contain $n_e \sim 1000$ neutral cells and what is measured — in terms of read counts — is the total size of the lineage. The effect of the beneficial cells will be clear within this lineage only when they comprise an appreciable fraction of the entire lineage i.e when $n(t) \sim n_e$. Therefore, instead of a waiting time of $1/s$ generations for the mutation to establish, detecting a mutation above the neutral cells takes a time of $(1/s)\ln(n_e s)$ generations. In terms of the occurrence time τ_{nut} of the mutation this places a limit

$$\tau_{mut} < t(\bar{x} = s) - (1/s)\ln(n_e s) \tag{76}$$

or in terms of the effect size

$$s \gtrsim \frac{1}{t - \tau_{mut}} \ln\left(\frac{n_e s}{c}\right) \tag{77}$$

in order to be detected as beneficial. Mutations must therefore occur earlier, by a factor of $\sim \ln(n_e s)$ than they do to simply establish. We note again that it is the effective lineage size n_e and not the bottleneck population size n_b that is relevant here because it was n_e that was used to infer τ_{est} from data. Equivalently, we could have replaced c by c/T and used the bottleneck size, $n_b = n_e/T$: as the times for establishment are typically multiple growth-dilution cycles, these are averaged over as accounted for either way.

To verify that the increase in mean fitness and finite lineage size does impose a fundamental lower limit on the detectability of small fitness effect mutations, we simulated an evolving population with parameters very similar to those in the experiment, namely a bottleneck population size of $N_b = 7 \times 10^7$, $L = 500,000$ barcoded lineages, and a distribution of mutations rates to different fitness effects which is uniform (see Section 12 for details). We do indeed observe (Figure 36) that for a given effect size s, there is a time limit before which mutations must occur in order to be detected. The limit imposed by clonal interference alone (large dashes) is the limit of what could possibly be detected even if every cell were tracked. The additional limit imposed by the finite initial lineage size is shown by the small dash line in Figure 36 which does indeed trace the line of mutations that are identified as adaptive in simulated data (for details of this data see Section 12).

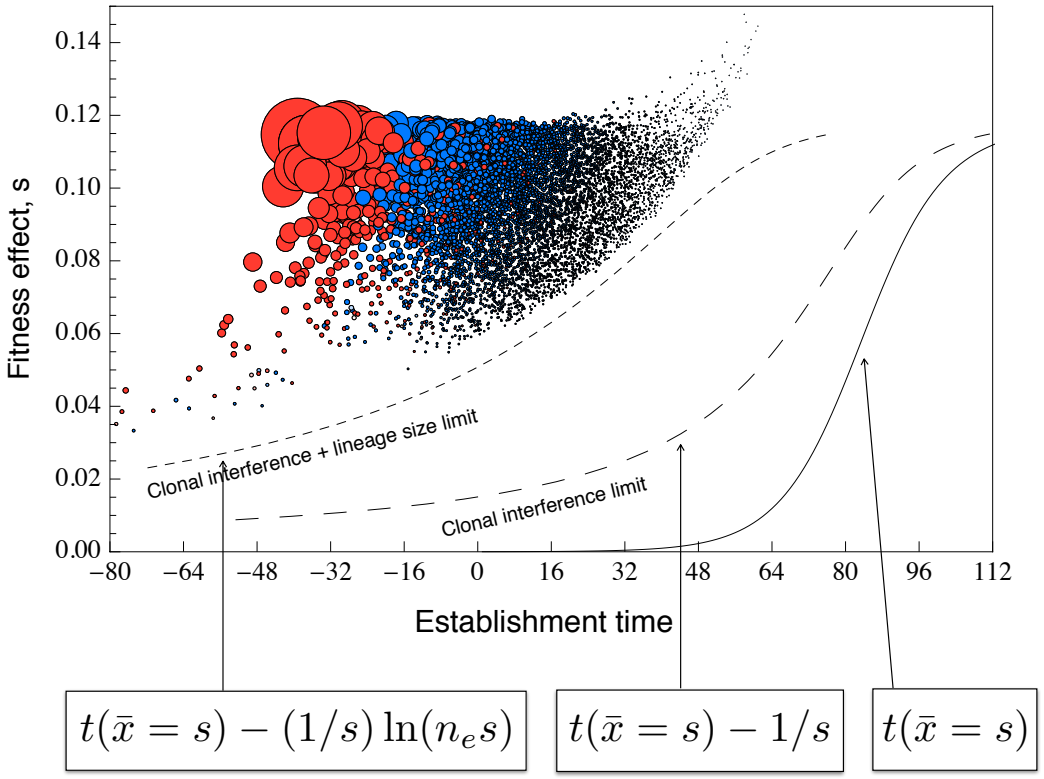


Figure 36: A plot of the fitness effect, s (y-axis) and establishment time, τ (x-axis) of all $\sim 8,000$ mutations detected in the simulated data set (described in detail in Section 12). Each data point is a beneficial mutation that arose during the simulation. The size (area) of the data point is proportional to the number of cells in which that beneficial mutation exists. Red points refer to mutations that are identified as adaptive in both replicate simulations and likely "pre-existing" while blue are mutations that were identified as adaptive in only one replicate (see Section 10). The solid line is that of the mean fitness increase defined by $t(\bar{x} = s)$. Only mutations occurring with $t_m < t(\bar{x} = s) - 1/s$ (larger dash line) can ever establish in the population. Only mutations occurring $t_m < t(\bar{x} = s) - (1/s)\ln(n_0 s)$ can both establish and be detected. The difference between the two dashed lines accounts for the additional time the mutation take to reach a size of $\sim n_0$ cells to be detected. Mutations that occur in the region between the two dashed lines can establish but remain undetected. Such mutations however never reach large sizes.

What does this imply for the experimental data? Putting in approximate numbers of $t \sim 100$ generations, $c \sim 3.5$, $n_0 \sim 10^3$, even if the mutation occurs immediately $t_m = 0$ it will only be detected if

$$s \gtrsim \frac{1}{100} \times \ln\left(\frac{1000 \times 0.04}{3.5}\right) \approx 0.026 \tag{78}$$

This agrees with the (s, τ) plots shown in the paper from E1 and E2 where mutations with $s < 3\%$ are rare.

It should be remembered that the limits (dashed lines in Figure 36) calculated here are approximate. In reality the time limit imposed on a detecting a mutation depends on the particular initial lineage size (not the median), on how the neutral cells fluctuate and during the expansion of the beneficial mutation, and on specific time it took for the beneficial mutation within the lineage to fluctuate up to high enough numbers to be detected (which can deviate from $(1/s)\ln(n_e s)$ by factors of $\pm 1/s$). These limits should therefore be interpreted not as strict boundaries but rather as lines that delineate a region where it becomes increasingly unlikely to detect a mutation.

9.3 Mutations with small fitness effects

Given the detectability limits imposed both by the evolutionary dynamics and by the initial size of lineages described above, what can we say about mutations of very small fitness effects? Are there mutations of small effect that remain undetected? Is the decline of mutation rates below the peak at $s = 4\%$ due to detectability issues, or because the rate is smaller to these weak mutations?

While it is hard to conclusively measure the rate of mutation to very small fitness effects, there are two important points relating to small fitness effect mutations. First, small fitness effects are inconsequential to adaptation unless they occur at implausibly high mutation rates. Second, although the precise shape of the distribution at $s < 4\%$ is difficult to measure, there is some tentative evidence suggesting that the peak observed in the inferred spectrum of mutation rates as a function of fitness effect really is a peak, and not just an artifact of detectability.

1. *Mutations of small fitness effect are inconsequential to the evolutionary dynamics we observe.*

   To impact the evolutionary dynamics of large populations, low fitness effect mutations have to occur at a very high rate. Using the deterministic approximation described in 11.1 the predominant s that drives the dynamics at time t is the s that maximizes

   $$\mu(s)e^{st} \tag{79}$$

   Given that the total rate to fitness effects in the $s \sim 4\%$ range is on the order of $U \approx 5 \times 10^{-5}$, what would the mutation rate to fitness effect $s < 2\%$ have to be, in order to drive the dynamics and be the "predominant" fitness class? This depends on what time we are interested in. A reasonable choice is the time when the mutants *as a whole* become an appreciable fraction of the population (they reach $\sim 10\%$ at $t \approx 70$ generations) which more properly corresponds to $t \approx 120$ generations of growth since there are 48 generations of growth before the separation of the replicates. From above, in order to have a similar impact as the $s = 4\%$ mutations, those with $s = 2\%$ or $s = 1\%$ would therefore have to occur at a rates of

   $$U(2\%) \approx U(4\%)\exp\left(100 \times (0.04 - 0.02)\right) \approx 5 \times 10^{-4} \tag{80}$$
   $$U(1\%) \approx U(4\%)\exp\left(100 \times (0.04 - 0.01)\right) \approx 2 \times 10^{-3} \tag{81}$$

   Given the per base pair mutation rate of 3×10^{-10}, these rates would require an implausibly large target size in excess of $\sim 10 - 50\%$ of the genome. Therefore it seems unlikely that mutations with small fitness effects can play a substantial role in the dynamics.

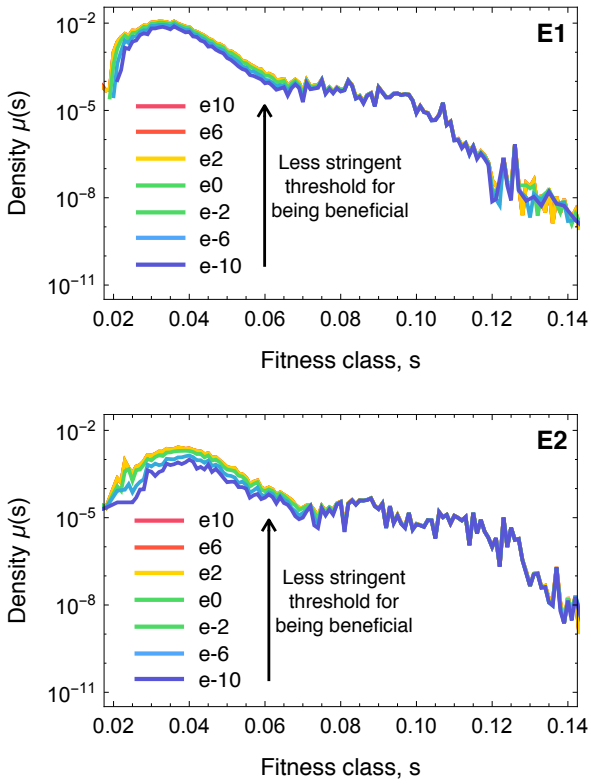


Figure 37: The mutation rate fitness spectrum that is inferred at varying adaptive thresholds. The number next to the colored lines is the threshold: any lineage whose posterior probability of the neutral hypothesis divided by the posterior probability of the beneficial hypothesis is smaller than the quoted threshold is identified as adaptive. For example the line with threshold of e^{-6} includes only lineages for which the beneficial hypothesis was $e^{6} \approx 400$ times more likely than the neutral hypothesis. Changing this threshold over 20-e-foldings does not significantly alter the inferred distribution. This insensitivity to the peak of the distribution suggests that there may indeed be a peak in the distribution at $s \approx 4\%$.

2. *Is the peak in the $\mu(s)$ distribution really a peak?*

   We observe a peak in the mutation rate fitness spectrum at $s \approx 4\%$. For fitness effect $s < 4\%$ it is difficult to conclusively say what the shape of the distribution is, or even whether it falls off or not. However do we have some tentative evidence that there are fewer mutations at fitnesses $s < 4\%$. If the decline in the rate of mutation to fitness effects $s < 4\%$ were due to detectability issues one would expect that as we change the threshold set for identifying lineages as either adaptive or neutral, this would significantly alter the observed spectrum at low fitnesses. Specifically, being more liberal in our calling of adaptive lineages should increase the number of small effect mutations we identify (as well as introducing more "false positives"). Figure 37 shows the inferred spectrum $\mu(s)$ for varying thresholds. The fact that the spectrum is so insensitive to varying the threshold (over a range of eight orders of magnitude) suggests that the mutation rate to effects $s < 4\%$ may indeed be smaller than it is at $s = 4\%$. In other words, the observed peak at $s = 4\%$ may really be a peak in the underlying spectrum.

9.4 The need for high frequency resolution.

- *How does the inferred distribution of fitness effects change with lower frequency resolution?*
- *What frequency resolution is required to fully observe the dynamics?*

To observe and measure properties of a mutation around establishment requires a frequency resolution capable of measuring lineages that are $\sim 1/s \sim 100$ cells in size. If the population size is N, then the frequency resolution required is $\sim 100/N$, which in our case is $\sim 10^{-6}$ meaning the number of barcodes needs to be $\sim 10^6$. (Note this will change slightly depending on the typical values of s in the system. Fitness effects on the percent scale are however typical for many evolution experiments e.g. [16])

This condition could, however, be considered overkill. If one does not require observing the mutation as it establishes, but instead only want to be sure that the expansion of a lineage is driven by a single beneficial mutation, one requires a lower frequency resolution: lineage sizes must be $\ll 1/U$, which in our case is $\sim 10^4$ cells per lineage. This condition would then require a frequency resolution of $\sim 10^4/N \sim 10^{-4}$ or in other words, $\sim 10^4$ lineages (though this clearly depends on the beneficial mutation rate of the system in question).

The frequency resolution range required to the observe evolutionary dynamics of beneficial mutations is therefore

$$\frac{1/s}{N} < f < \frac{1/U}{N} \tag{82}$$

where s is the typical selective effect and U the total mutation rate. The lower limit enables one to observe all established mutations just as they reach establishment frequencies in the population. The higher limit does not permit this, but does enable one to observe and measure the expansion of a lineage due to a single beneficial mutation.

To illustrate this point, Figure 38 shows what the inferred $\mu(s)$ would be with progressively lower frequency resolution. As one might guess, having the ability to measure only frequencies in the percent range means one misses the vast majority of mutations. A related figure is shown in the main paper where we plot how many adaptive lineages would be identified above a given fitness, for varying frequency resolution. These plots confirm that to fully observe the evolutionary dynamics requires a frequency resolution $\sim 10^{-5}$.

Figure 4 of the main text shows another way of seeing the different distribution of fitness effects one would see at lower frequency resolution. To make this plot we collected all barcodes that reach a frequency above the quoted threshold by $t = 100$ (based on their inferred (s, τ) values), and plotted the distribution of their inferred fitnesses, s. At lower frequency resolution one is restricted to observing far fewer beneficial mutations, and those that are observed are confined to a narrow fitness range.

Required sequencing depth. A related question to the above is: what sequencing depth at each time point is required in order to fully characterize the dynamics of each lineage? If the above considerations require a number of unique lineages on the order of $L \sim 10^6$, then the total read depth required to measure each of these well will be ~ 10 reads per barcode (depending on the fluctuations one is willing to accept) which results in $\sim 10^7$ reads per time point. In our case, read depth across most time points was between $2 - 8 \times 10^7$ (See Table 2). The exact depth required depends on how much variance in frequency one wants to tolerate. The rule of thumb we used for our data was that the read depth in a lineage should introduce about the same amount of variance in frequency as other factors such as drift, variance in offspring number and variance introduced via DNA extraction and amplification (See Section 5). This means picking a read coverage that is roughly equal to the bottleneck population size (in our case 7×10^7 cells), so for most purposes

$$R \sim N_b \tag{83}$$

It should be noted however that a deeper read coverage could, in some contexts, be useful. For example, if one wanted to precisely characterize the distribution of offspring number (see Section 5.4) one would want the variance introduced by sampling of reads at the sequencer to be lower than the other contributions so one would want a depth $R \gg N_b$.

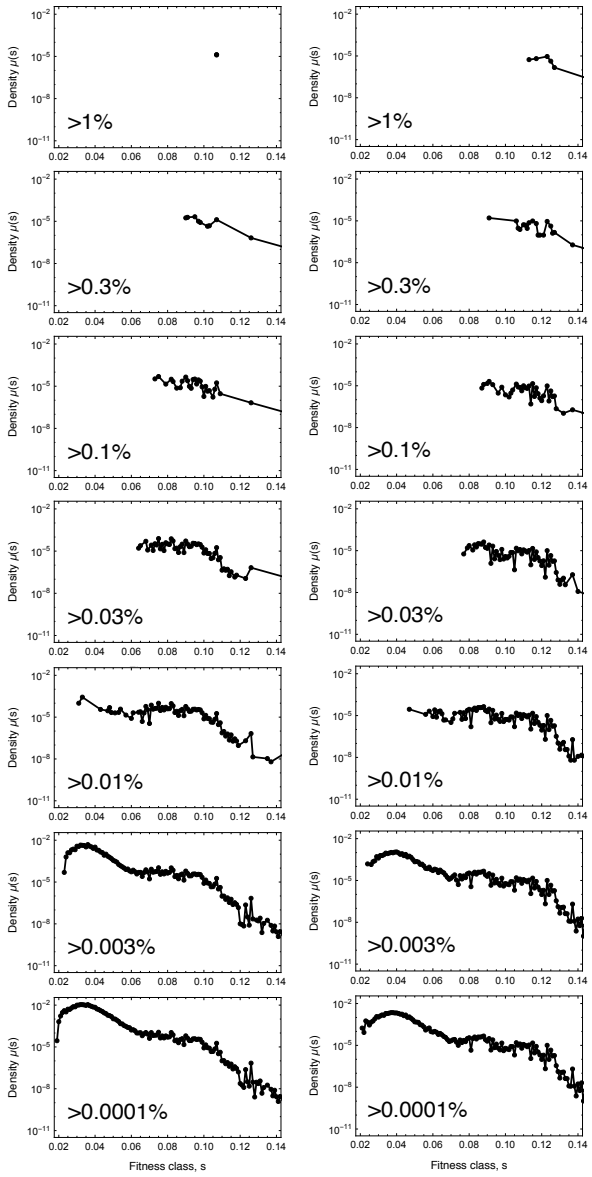


Figure 38: The inferred mutation rate fitness spectrum for E1 (left) and E2 (right) that would have been observed at different frequency resolutions. For example, the top panel shows the distribution observed if only lineages that reach $> 1\%$ in frequency (by $t = 100$) could be identified. A frequency resolution of $O(10^{-5})$ is required to fully characterize the distribution and the evolutionary dynamics.

10 Pre-existing mutations

Mutations inevitably arise during the period of common growth before the beginning of growth-bottleneck cycles, i.e. $t < 0$ (Figure 39). Here we estimate how many beneficial mutations occur during this period of prior growth and calculate how many of these are then sampled into, and establish, in both replicates and how many are sampled into and establish in one replicate but not the other. We observe $\sim 6,000$ lineages that are identified as adaptive in *both* replicates E1 and E2. The fitness effects of these mutations, measured in the glucose limiting liquid environment used for $t > 0$, are almost exclusively in the range $2.5\% < s < 4.5\%$ and have early establishment times $-100 < \tau < -20$ (Figure 40). Together with estimates of how many mutations could have arisen independently (Section 10.3), these facts leads us to conclude that $\sim 5,500$ lineages must have accumulated beneficial mutations that likely arose in the period

of growth prior to the insertion of barcodes (Regions 1 and 2 in Figure 39) which were sampled into, and established in, both replicates. In the following we outline the arguments that lead to these conclusions. There are two distinct environments prior to $t = 0$: growth in YPD (regions 1 and 2 in Figure 39) and growth on SC - ura plates (region 3 in Figure 39). We consider these separately as it is likely mutations have different fitness effects in the two environments.

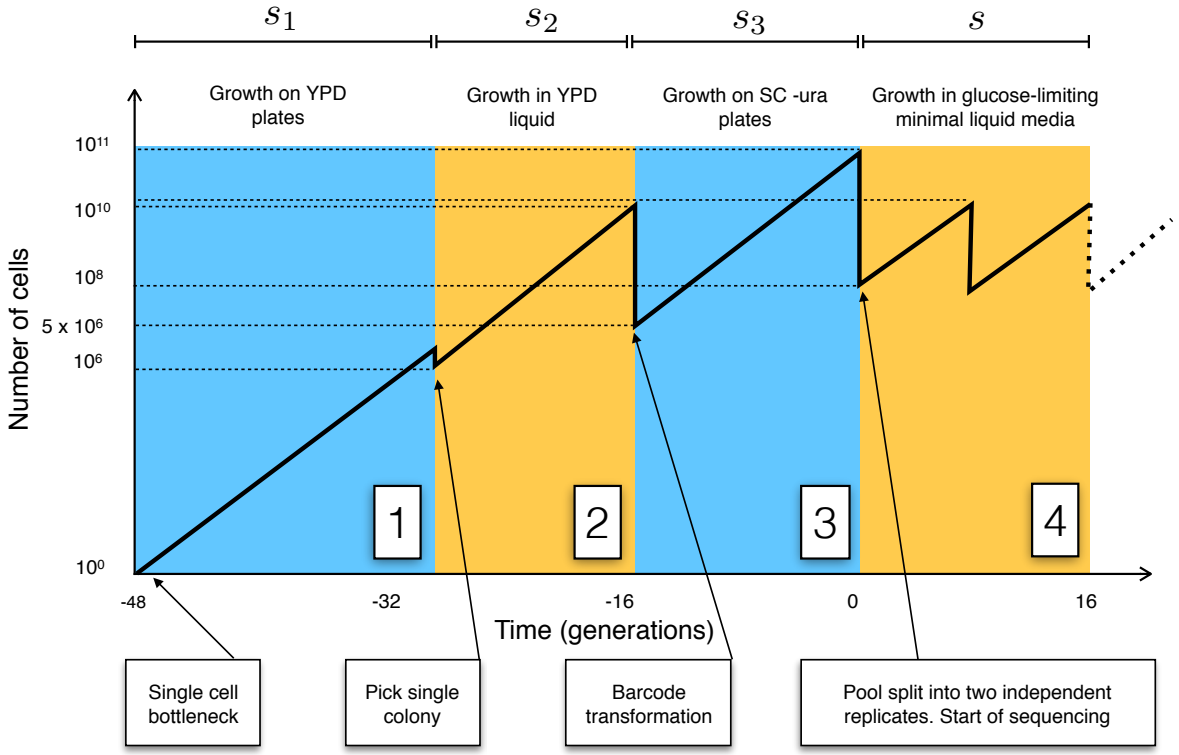


Figure 39: A single cell is grown on a YPD+Kan selectable plate to $\sim 10^6$ cells which equates to $T \sim 20$ generations of growth (Region 1). Most of this colony $\sim 10^6$ cells is taken and further grown up though ~ 13 generations in YPD liquid media to a total of $\sim 10^{10}$ cells (Region 2). These cells are then used in the transformation reaction to incorporate the DNA barcodes. Of the $\sim 10^{10}$ cells that begin this process, only $\sim 5 \times 10^6$ cells successfully incorporate a barcode. Therefore each barcode tag is incorporated independently into ~ 10 different cells. These grow through a further $\sim 16 - 18$ generations (Region 3). Although more mutations arise in region 3 than in regions 1 and 2, most of these are present in a small number of cells and are unlikely to be sampled into the replicates E1 and E2 (Region 4). We observe $\sim 6,000$ lineages that are adaptive in both E1 and E2. Of these, $\sim 5,500$ occur in regions 1 and 2 and ~ 500 are mutations that occurred independently in the same lineage in region 4. Very few (perhaps ~ 200) lineages from region 3 establish in both replicates.

10.1 Pre-existing mutations from growth in Regions 1 and 2, before barcoding

We begin by determining how many mutations accumulate in the cells prior to barcoding. Growth before the barcodes are inserted is in YPD media. The population size of all cells grows approximately as:

$$n(t) \approx 2^t \tag{84}$$

for $0 < t < T$ with $T \approx 33$ (we ignore the very slight bottleneck between regions 1 and 2). Suppose that the fitness effect of the pre-existing mutations in YPD is s_{ypd} and that the feeding population gives rise to these mutations at a rate u_b. The first mutation will enter when $u_b 2^t \sim 1$ and reach a size $u_b 2^T e^{s_{ypd} T}$ by the end of the growth. It is unlikely that the factor of $e^{s_{ypd} T}$ is very large since the fitness effects are

typically $\sim 3-5\%$ and the time of growth is only ~ 33 generations. The second mutant to enter will do so when $u_b 2^t \sim 2$ and reach size $\sim 1/2$ as large as the first. The third will be $1/3$ as large as the first and so on. Using results from the exponential feeding process in Section 13.5, (here $\alpha \sim 1$), the mutations that enter during this period have typical sizes given by the following series

$$\text{(Total number of beneficial mutant cells)} \approx u_b \times 2^T e^{s_{ypd}T} \times \left(1 + \frac{1}{2} + \frac{1}{3} ... + \frac{1}{m}\right) \tag{85}$$

where each of the terms in the sum is the typical size of the first, second, third and eventually, mth mutation that enters and where $m \approx u_b 2^T$ is the total number of unique mutations that likely enter during this period of growth. The total number of beneficial cells is then the sum of the above, which is approximately

$$\text{(Total number of beneficial mutant cells)} \approx u_b 2^T e^{s_{ypd}T} \log(u_b 2^T) \tag{86}$$

Since the final population size is large enough that $u_b 2^T \gg 1$, the total number of mutations that enter during this period, m, is large and the sum is not typically dominated by the earliest mutations. (Note, however, that the size of the first mutation population has a power-law tail that decays as $1/n^2$ out to $2^T e^{s_{ypd}T}$ due to the possibility of an anomalously early mutation occurring, though such large fluctuations are unlikely.) After this initial growth, a bottleneck is performed (corresponding to the 5×10^6 cells that successfully incorporate a barcode) which reduces the number of beneficial cells to approximately

$$\text{(Total number of beneficial mutant cells after barcodes inserted)} \sim 5 \times 10^6 u_b e^{s_{ypd}T} \ln(10^{10} u_b) \tag{87}$$

How many lineages are called as adaptive because of these cells? Assuming that the fitness effect of pre-existing mutations is similar in YPD as in the growth media of the experiment i.e. $s_{ypd} \approx s \approx 3.5\%$, the additional factor of $e^{s_{ypd}t}$ is order e. If one assumes a reasonable mutation rate of $u_b \sim 5 \times 10^{-5}$ (consistent with our later inferences) this would predict that $\sim 6,000$ cells with beneficial mutations from regions 1 and 2 should be sampled into each of the replicates. Each of these beneficial cells will likely have received a unique barcode, since the number of barcodes ($\sim 500,000$) is much larger than the number of beneficial cells that make it through the bottleneck. Note that for this analysis is does not matter much whether *each* individual mutation has the same fitness in YPD and the main experimental medium: as long as the distribution of fitnesses is similar at level of a few % — including later-beneficial ones being initially weakly deleterious — the behavior will not be much different because the product sT is never substantially larger than one.

How many of these $\sim 6,000$ mutations will be establish and be detected in both replicates? Here an important point to consider is that each barcode is incorporated ~ 10 times independently. If one of these ~ 10 cells has a beneficial mutation then, at the start of the experiment the number of cells sharing this beneficial mutation will be $\sim 10-30$ (at the bottleneck) or $\sim 100-300$ cells using the effective population size. Because of this, almost all of the $\sim 6,000$ lineages with beneficial mutations from regions 1 and 2 will establish in both replicates E1 and E2, and will have early establishment times. In fact the establishment times can be estimated using the fact that 10% - 30% of the lineage will have beneficial cells so

$$n(t) \approx 0.3 n_e e^{st} = \frac{e^{s(t-\tau)}}{s} \qquad \Rightarrow \qquad \tau \approx -\frac{1}{s} \ln(0.3 n_e s) \sim -75 \text{ generations} \tag{88}$$

For $s \sim 4\%$. Figure 40 shows that most mutations that are adaptive in both replicates E1 and E2 have substantially early establishment times: in the region expected if they were to have arisen in regions 1 and 2.

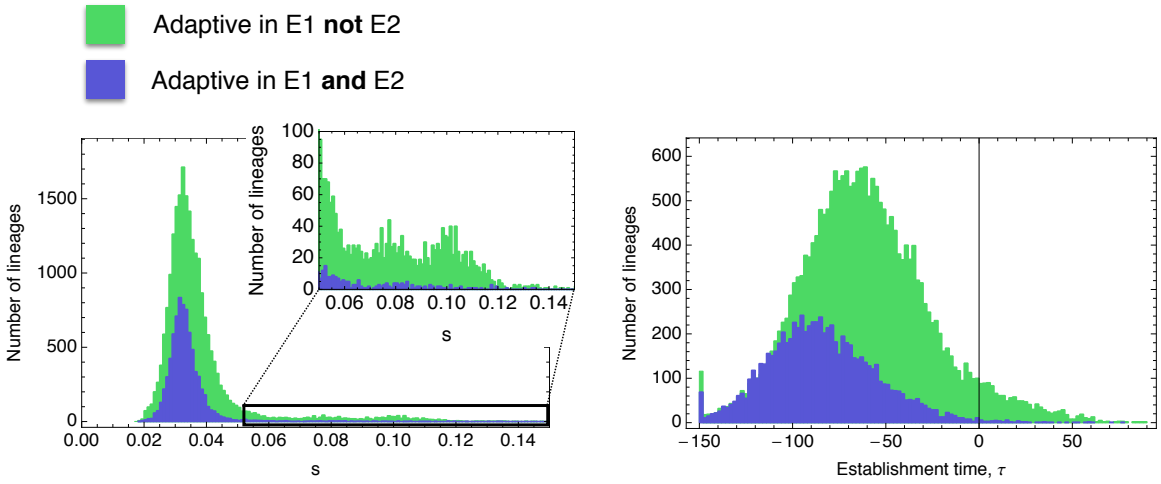


Figure 40: Fitness effect (left) and Establishment time (right) distributions of lineages colored accord to whether it was adaptive in both replicates ($\sim 6,000$ lineages, purple) or only in E1 ($\sim 20,000$ lineages, green). Lineage adaptive in both replicates have fitness effects in the $2.5\% < s < 4.5\%$ range. Their establishment times are early: in the $-150 < s < -50$ range.

10.2 Pre-existing mutations from growth in Region 3, after barcoding

Region 3 in Figure 39 is growth on SC -ura plates after barcoding but before the separation of the replicates E1 and E2. The $\sim 5 \times 10^6$ cells that incorporate a barcode grow into colonies, which takes $T \sim 16 - 18$ generations. While the population grows from $5 \times 10^6 \to 10^{12}$ cells further beneficial mutations accumulate. Denoting the fitness effect of the pre-existing mutations as s_{sc} in this environment, each of the mutations that enter in this region will be of typical size:

$$2 \times 10^{12} e^{s_{sc} T} u_b \left(1 + \frac{1}{2} + \frac{1}{3} \ldots \frac{1}{m}\right) \tag{89}$$

where again each term is the size of the first, second mutation and so on and where $m \sim 2 \times 10^{12} u_b$ is the total number of mutations that occur during region 3. While this number is likely substantial, most of the mutations exist in small numbers of cells and hence are unlikely to be sampled through the bottleneck of $\sim 10^8$ into both replicates E1 and E2.

Number of mutations from Region 3 that establish in both E1 and E2. The mean size of the mutations after the bottleneck will be

$$10^8 e^{s_{sc} T} u_b \left(1 + \frac{1}{2} + \frac{1}{3} \ldots \frac{1}{m}\right) \tag{90}$$

If the number of cells of a particular mutation sampled into E1 and E2 are n_1 and n_2 respectively, then the probability that they establish in both replicates is $\approx (1 - e^{-n_1 s_{sc}/c})(1 - e^{-n_2 s_{sc}/c})$. If $k \in [1, m]$ enumerates all the mutants in Eqn. 90, then the expected number that establish in both E1 and E2 is

$$\sum_k^m \left[\sum_{n_1, n_2} \left(1 - e^{-n_2 s_{sc}/c}\right) \left(1 - e^{-n_1 s_{sc}/c}\right) P(n_1) P(n_2) \right]_k \tag{91}$$

Mutations that have a mean size after bottleneck from Eqn. 90 of $\langle n_k \rangle \gg c/s_{sc}$ all establish in both because $1 - e^{-n_1 s_{sc}/c} \sim 1$. For $\langle n_k \rangle \ll 1$ the probability of establishing in both is $\sim s_{sc}^2 \langle n_k \rangle^2$ (since $P(n_k > 1) \ll P(n_k = 1)$ one can consider only the possibility of singletons) meaning the expected number

of rare mutants that establish in both replicates is

$$10^8 e^{s_{sc}T} u_b \sum_k^m \left(\frac{s_{sc}/c}{k}\right)^2 \tag{92}$$

which is dominated by the k for which $\langle n_k \rangle \sim c/s_{sc}$. The expected number that establish in both is therefore dominated by how many mutants get sampled with sizes of order $\sim c/s$ and above. This is the same as asking how many terms in the series in Eqn. 90 are of order c/s which gives

$$(\#\text{ mutations that establish in E1 and E2 from region 3}) = 10^8 e^{s_{sc}T} u_b (s/c) \tag{93}$$

Number of mutations from Region 3 that establish in E1 or E2. The expected number of mutations that establish in one replicate and not the other is

$$2 \sum_k \left[\sum_{n_1,n_2} e^{-n_1 s/c} \left(1 - e^{-n_2 s/c}\right) P(n_1) P(n_2) \right]_k \tag{94}$$

In contrast to the above case where mutations establishing in both was dominated by those that have expected sizes of $\sim c/s$, for mutations that establish in one and not the other the dominant contribution to the sum for the number of mutants come from mutations that are present in essentially one copy after bottlenecking, the vast majority of all the mutations that occur:

$$10^8 e^{s_{sc}T} u_b \sum_k^m \frac{(s/c)}{k} \sim 10^8 e^{s_{sc}T} u_b (s/c) \ln(m) \sim 10^8 e^{s_{sc}T} u_b (s/c) \ln\left(10^{12} u_b\right) \tag{95}$$

Assuming the mutation rate to the pre-existing mutations is similar to the value inferred for regions 1 and 2 ($\sim 5 \times 10^{-5}$) and that the fitness effect of pre-existing mutations in SC -ura plates is (s_{sc}) is similar to that in both YPD and in the glucose limited media, this predicts that a further ~ 200 lineages would be adaptive in *both* E1 and E2 because of beneficial mutations from region 3 and $\sim 4,000$ lineages are adaptive in one replicate but not the other.

The $\sim 4,000$ beneficial mutations that are predicted to establish in one replicate but no the other do not have substantially negative establishment times. The majority of these mutations begin as singletons in one of the replicates and hence the distribution of their establishment times will be similar to that of a mutation present in a single copy at $t = 0$. Establishment times of these mutations can therefore extend back to as early as $\tau \sim -1/s$ generations but not substantially further and likely (in part) account for the $\sim 5,000$ mutations whose establishment times are $-50 < \tau < 0$.

10.3 Number of lineages with adaptive mutations in both replicates if acquired independently

We observe $\sim 6,000$ mutations that are identified as adaptive in both replicates E1 and E2, which broadly agrees with how many we would expect as a consequence of common growth prior to barcoding. However, barcodes that are called as adaptive in both replicates E1 and E2 could also arise by truly independent mutations occurring in the same barcode at times $t > 0$. How many barcodes are expected to be called as adaptive in both E1 and E2 because of independently occurring mutations? Assuming all lineages have the same probability of accumulating a beneficial mutation we would predict

$$\#\text{ independent mutations adaptive in both} \sim \frac{(\#\text{adaptive in E1}) \times (\#\text{adaptive in E2})}{\text{Total number of lineages}} \sim 500. \tag{96}$$

(This is not quite right because not all lineages have the same probability of accumulating a beneficial mutation. Larger lineages have a greater chance than small ones. If one accounts for the fact that the probability of accumulating a beneficial mutation is not uniform because there is an initial distribution

of lineage sizes this answer gets modified by a factor $\langle n_0^2\rangle/\langle n_0\rangle^2$ where n_0 is the initial lineage size and averages are over the distribution of initial lineage sizes. However since the distribution of initial lineage sizes is almost exponential (Figure 14) this factor is ~ 2 and therefore the number of lineages we expect to be identified as adaptive due to independent mutations occurring in the same barcode remains < 1000). This leaves $\sim 5,000$ mutations that must be pre-existing mutations (occurred with $\tau_{mut} < 0$) that were sampled and then established in both replicates. This number agrees with the number we expect from the previous estimates.

10.4 Checking self-consistency using the high-fitness mutations

Another check on the consistency of our inferences comes from considering the high fitness effect mutations. We infer that high fitness-effect mutations ($s > 0.08$) occur with a rate $U_{s>0.08} \sim 2 \times 10^{-7}$. If this is the case, then we should expect a number of these to occur during the common growth before barcoding and be sampled into both E1 and E2. How many should be sampled into both? The earlier estimates for mutations in the $3-5\%$ range assumed a mutation rate of $U_{0.03<s0.05} \sim 5\times 10^{-5}$ and concluded that $\sim 6,000$ mutations should be common across replicates. If mutation rates are reduced to $U_{s>0.08} \sim 2\times 10^{-7}$ this would predict

$$\text{Lineages with } s > 0.08 \text{ adaptive in E1 and E2} \approx 0.004 \times 6000 \approx 25 \tag{97}$$

This broadly agrees with the number we see (29).

10.5 Identifying pre-existing mutations

For any given adaptive lineage can we say whether the mutation within it was pre-existing? There are two pieces of information that inform us: whether the same lineage is identified as adaptive in both replicates and how early the establishment time of the mutation is.

If a mutation is identified as adaptive in both replicates E1 and E2 then it is likely it was pre-existing since only ~ 500 are expected by chance to occur independently in the same lineage. In the $s-\tau$ scatter plots in the paper these are colored purple. As discussed in Sections 10.1 and 10.2, the majority of these mutations likely arose in regions 1 and 2 during the common growth before barcoding (Figure 39).

Some mutations however will have occured during prior growth but will establish in one replicate and not the other. The majority of these mutations arose during growth after barcoding, before the splitting of the replicates (Region 3, Figure 39). Furthermore, most of the beneficial mutations that accumulate in Region 3 and that get sampled into one of the replicates are most likely to do so in a single copy (as discussed above in Section 10.2). It is not possible to individually distinguish such mutations from mutations that arose in the first few generations of growth for $t > 0$. The reason is that the distribution of establishment times for a mutation present in a single cell at $t = 0$ is broadly distributed around $t = 0$ with errors of $\pm 1/s$ meaning that the distribution of establishment times for these mutations overlap substantially.

However, while we cannot individually distinguish mutations that arose in the region $-16 < t < 0$ from those that arose in the first few generations $t > 0$, we can account for their effect on the total number of cells and hence for their effect on the total beneficial mutation rate.

If mutations in the fitness range $[s, s+\delta s]$ occur at rate $\mu(s)\delta s$, using the same arguments as outlined in Section 10.2, the total fraction that accumulate in Region 3 (Figure 39) and that make it through the bottleneck to be sampled into one of the replicates is

$$\text{(fraction of cells from region 3)} = \mu(s)\delta s \ln(N_f \mu(s)\delta s) \tag{98}$$

where $N_f \approx 10^{12}$ is the total number of cells at the end of region 3 before the bottleneck is imposed. As we outline in the following section, this fraction of cells that can be assigned to mutations entering in region 3 can be taken into account when estimating the beneficial mutation rate to each fitness effect.

11 Inferring the mutational fitness spectrum $\mu(s)$

The spectrum of the fitness-dependence of mutation rates, $\mu(s)ds$, we define as the mutation rate per cell per generation to mutations whose fitness effects are in the range $[s, s+ds]$. We can use the inferred values of (s, τ) for each barcode lineage to infer $\mu(s)ds$ for beneficial mutations in two distinct, though related ways described in the following two sections.

Separating out pre-existing mutations. As highlighted in Section 10, evolution does not begin at $t = 0$ but rather the moment from the last common ancestor. Mutations can, and do, arise in the generations of growth preceding the separation of the two replicates E1 and E2. This prior growth passed through a number of different environments e.g. growth in liquid and growth on selectable plates. To determine the distribution of fitness effects $\mu(s)ds$ in the constant environment at $t > 0$ therefore requires the ability to account for how many mutations we observe come from the period of prior growth. We outline how we do this in each of the following two sections.

The method we actually use to infer the distribution of mutation rates across fitness effects quoted in the main paper is the deterministic approximation (described in the next section). We note however that both methods give broadly consistent conclusions.

11.1 Deterministic approximation, "Predominant" s, and stochastic transition

The deterministic approximation makes use of the fact that, if mutations of effect sizes in the range $[s, s+ds]$ are being fed from a large population of N_e neutral cells at a rate of $\mu(s)ds$, then, provided the product $N_e\mu(s)ds \gg 1$ the total fraction of cells in the population with fitness in the range $[s, s+ds]$ is

$$f(ds, t) = \frac{\mu(s)ds}{s} e^{st} \tag{99}$$

Therefore by measuring $f(ds, t)$ we can infer the mutation rate to that range of fitness effects. In order to do this correctly however, we must account for the effects of pre-existing mutations.

Accounting for pre-existing mutations. The inevitable expansion of mutations as the population grows up from a single cell to the final barcoded library (i.e. those that arose prior to $t = 0$) affect estimates of the beneficial mutation rate using the deterministic approximation, because they increase the fraction $f(ds, t)$ of cells in a given fitness range. How do we account for this? First, any lineage that is adaptive across both replicates E1 and E2 is excluded from the analysis of rates: this excludes the mutations that very likely arose before the barcoding process (regions 1 and 2 in Figure 39). Second we calculate what fraction of cells in a given fitness range likely arose after barcoding but before the separation of the replicates (region 3 in Figure 39).

$$\text{(additional fraction from mutations arising in region 3)} = \mu(s)\delta s \ln(N_f \mu(s)\delta s) e^{st} \tag{100}$$

Where $N_f \sim 10^{12}$ is the maximum population size the population reached after barcoding, before the separation of the replicates (see Figure 39). Accounting for this additional fraction of cells means that the expression for $\mu(s)\delta s$ becomes

$$\boxed{\mu(s)ds = \frac{f(ds, t)}{1 + s\ln(N_f \mu(s)\delta s)}} \tag{101}$$

which is the formula we use to infer the distribution of mutation rates. We note that the magnitude of the logarithmic term from mutations that arose after barcoding but before the $t = 0$ bottleneck assumes that over the $T \sim 16$ generations the exponential growth advantage of the mutants is not a huge effect since $e^{sT} - 1 \sim 2$. For the typical number we infer here $\mu(s)\delta s$ in the range $[3\%, 5\%]$ is $\sim 10^{-5}$ and $N_f \approx 10^{12}$ meaning that the logarithmic term is half the size of the original term, that is, the contribution to total number of cells as a function of time from mutations that arose in the time range $-15 < t < 0$ is about half of that from those that arise $t > 0$.

Predominant s. The deterministic approximation, which holds provided $N\mu(s)ds \gg 1$,can be used to determine which range of s contributes most to driving the mean fitness at a given time. The major contributor to the mean fitness is the class of mutation that has the most cells, i.e. which s mutants are most abundant in the population. In the deterministic approximation this is determined solely by the product

$$\mu(s)e^{st} \tag{102}$$

(provided $st \gg 1$). Maximizing this over s we can determine the predominant fitness range $s_{dom}(t)$ that is most abundant at time t. This $s_{dom}(t)$ must satisfy

$$\frac{d\mu}{ds}|s_{dom} = -t\mu(s_{dom}) \tag{103}$$

(if there are multiple solutions then the one with the largest $\mu(s_{dom})e^{s_{dom}t}$ is the predominant one). One can visualize this by drawing the distribution of fitness effects on a log-scale and then constructing a line with gradient $-t$. Lowering the line with gradient $-t$ from above, the first point on the $\ln(\mu(s))$ curve that is tangent to the line determines the predominant s, as shown in Figure 41.

Break-down of the deterministic approximation. Eventually, as rarer mutations enter and expand, the predominant fitness s_{dom} increases enough that one enters a regime where

$$N_e\tilde{U}(s > s_{dom}) \sim 1 \tag{104}$$

that is, where the predominant mutations are rare enough that only one (or perhaps a handful) will contribute to the expansion. In this case the dynamics of the mean fitness and of the expansion of the fittest class of mutations in the population becomes stochastic because they are no longer the aggregate effect of a large number of independent mutations. This is beginning to be the case for mutations with $s > 0.08$ in our data. Consider mutations in the range $0.095 < s < 0.105$. The rates of beneficial mutations inferred here are $\sim 10^{-7}$, meaning we expect only $N_e \times 10^{-7} \approx 50$ mutations to contribute to the expansion. Even further out, in the range $0.13 < s < 0.14$ mutation rates are inferred to be $\sim 10^{-9}$ giving a $N_e \times 10^{-9} \approx 0.5$. Such rare mutations will be stochastic, and indeed one finds that the two replicates look significantly different and noisy in this region (see Figure 3 in main paper) which contrasts with how similar they look in the more deterministic region at earlier times. When the deterministic approximation breaks down, reliably inferring rates becomes more challenging as only a small number of events ever occur. In addition, the details of the growth process and the variance in offspring number now become important because it is the first mutational event rather than the mean of the effects of many that matters most.

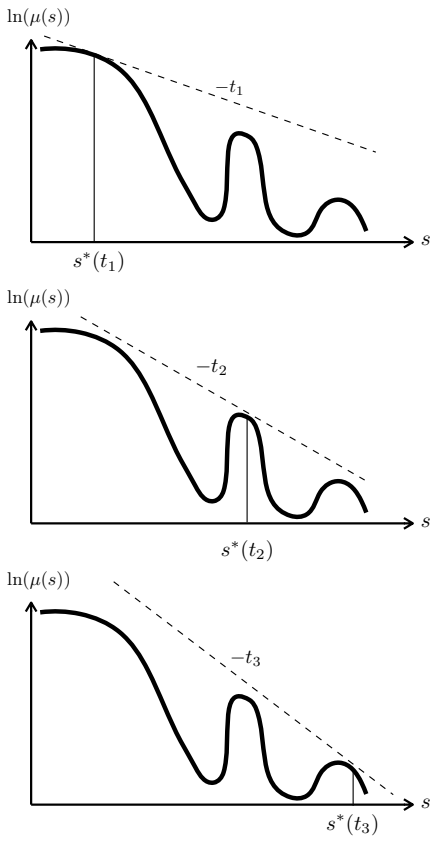


Figure 41: The fitness effect s that drives the mean fitness higher depends on the both the shape of the mutational fitness spectrum $\mu(s)$ and on the time t. The highest point on the curve $\ln(\mu(s))$ that is tangent to the line $-t$ defines the fitness effect s that is dominating the mean fitness rise (the class with the most cells). If the shape of $\mu(s)$ is non-convex (as we measure in this experiment) then the s_{dom} than drives the dynamics takes discontinuous jumps.

11.2 Estimating Errors on $\mu(s)$ from the deterministic approximation

The deterministic approximation for inferring $\mu(s)$ only has significant errors when the number of mutations contributing to the growth of a fitness class is small. More concretely, the size of a fitness class is

$$n(t) = \nu \frac{e^{st} - 1}{s} \tag{105}$$

where $\langle \nu \rangle = N\mu(s)\delta s$ and its variance is $\sim N\mu(s)\delta s$. This means that the typical error in the inferred values of $\mu(s)\delta s$ are order $\pm\sqrt{\mu(s)\delta s/N}$. These are shown as shaded regions in Figure 3 of the main text.

However, there are also other sources of error. In our case there is uncertainty on which lineages accumulated mutations before or after the separation of the two replicates at $t = 0$. In order to capture the uncertainties associated with this, we used a "conservative" and "liberal" approach to estimate how many pre-existing barcodes there are. The conservative approach estimated the number by conditioning on barcodes being adaptive across both replicates and having establishment times $\tau_{est} < -2/s$. The liberal approach approach did not condition on establishment time. In the distributions of $\mu(s)$ in Figure 3 of the main text, the upper bound for $\mu(s)$ is calculated using the conservative (i.e. smaller) set of pre-existing mutations excluded combined with the $+\sqrt{\mu(s)\delta s/N}$ uncertainty from above, while the lower bound is calculated using the liberal (i.e. larger) set of pre-existing mutations excluded and the $-\sqrt{\mu(s)\delta s/N}$ uncertainty from above.

11.3 Comparison of $\mu(s)$ from E1 and E2

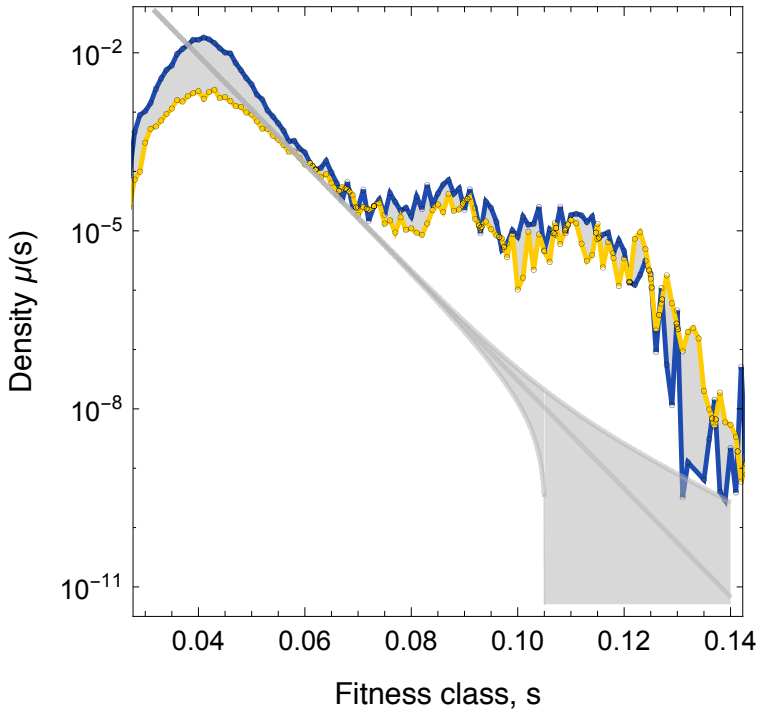


Figure 42: Comparing the inferred $\mu(s)$ distribution from E1 (blue) and E2 (yellow). Accounting for the $\delta s = 0.008$ systematic difference between E1 and E2, the inferred distributions are strikingly similar sharing features at low, intermediate and high fitness. Neither of the two distributions is consistent with an exponential distribution (gray line with errors as gray shading). The errors on the exponential are calculated based on errors associated with the small numbers of mutations in a fitness class (see Section 11.2)

11.4 Inferring $\mu(s)$ by counting the number of mutations in δs

Another method to infer the mutation rate to the range $[s, s + \delta s]$ is to count the number of mutations that have been identified as adaptive over a period of t generations. This method does not weight early (and hence abundant) mutations any more than late ones unlike the deterministic approximation. This has the added benefit that pre-existing mutations that arose after barcoding, but before the separation of the replicates (region 3, Figure 39) have less of an impact on estimates. However this method is more sensitive to our confidence in calling adaptive lineages in the first place, and because it counts number of mutations versus numbers of cells is sensitive to details of the growth process e.g. the changing population size and the variance in offspring number. Because of this, we elected not to use this method to infer the distribution of mutation rates across fitness effects quoted in the paper, however we did infer the distribution using this method to check our estimates were broadly consistent with the deterministic approximation method. We outline below how we did this.

If the effective population size governing the rate at which beneficial mutations establish is N_e then after t generations the number of mutations that have entered and established should be

$$(\text{number of mutations in } \delta s) \approx N \times (\mu(s)\delta s) \times (s/c) \times t \tag{106}$$

where the term (s/c) is the probability that a mutation establishes in the population given it has entered. For our purposes this simple expression must be modified to account for two things:

1. The feeding population N_e declines over time because of the mean fitness increase

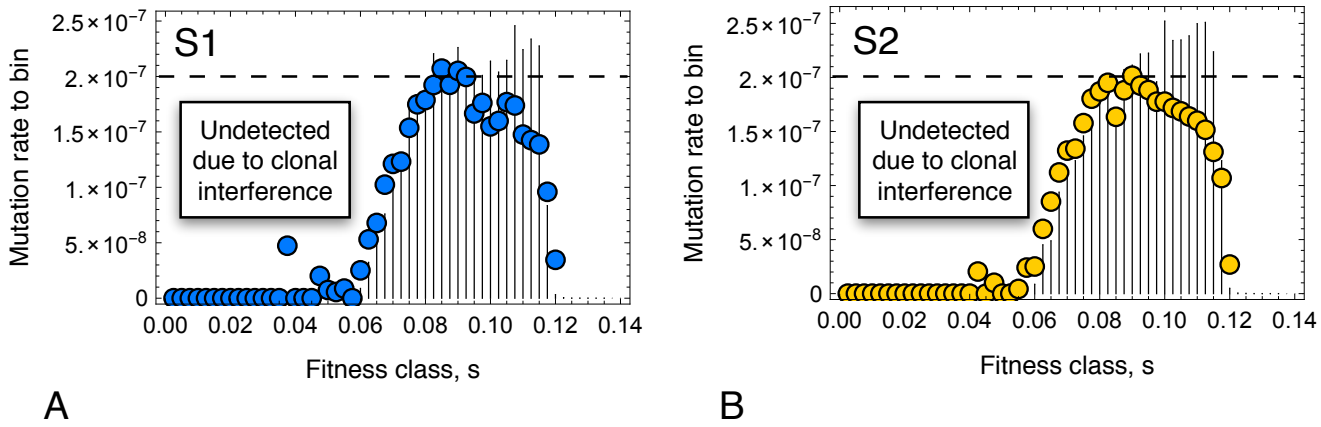


Figure 43: Comparing two methods of inferring the distribution of mutation rates to different fitness effects from simulations. The simulations were performed using a uniform distribution of mutation rates (shown as the horizontal dashed line). Each plot shows the inferred rates to bins of width $ds = 0.0025$ for the deterministic approximation (black lines) and method using raw counts of mutations (circles). (A) shows the inferences for S1 while B shows the inferences for S2. Clonal interference and finite lineage size prohibit small effect mutations from ever being detected. The reason for the consistent decline of the circles at later times is an artifact caused by estimations of the time window available for mutations to occur. As shown in Figure 46 high fitness effects that approach the maximum are estimated to have longer to occur than they actually do.

2. The relevant time determining the number of mutations detected is not the time of the evolution t but the time $t - (1/s)\ln(n_e s/c)$ (see Figure 36) which is the time before which a mutation must occur to have a significant chance of being detected.

With these two modifications the number of mutations identified as adaptive in the range $[s, s+\delta s]$ becomes

$$(\text{number of mutations in } \delta s) = (\mu(s)\delta s) \times (s/c) \times N_e \int_0^{t-(1/s)\ln(n_0 s/c)} e^{-\bar{x}(t)} dt \tag{107}$$

which can be inverted to estimate $\mu(s)$. We verified that this method produced similar values for the mutation rate as the deterministic approximation and that both produce inferences in agreement with known rate of beneficial mutation from simulations (Figure 43). This method is likely a better approach for inferring the mutation rates to higher-fitness effect mutations that occur less often since it weights mutational events evenly. It is also less likely to suffer from distinguishing between pre-existing mutations and those that arise after the separation of the replicates.

12 Simulated data set

To test our methods we analyzed a simulated data set that recreates as faithfully as we can all aspects of the true data set — including the 48 generations of common growth prior to growth-bottleneck cycles — and analyzed it using the same methods as for the real data. As in the real data we "evolve" two replicate simulations S1 and S2 that are seeded from a common pool of barcodes that has undergone the prior growth. By comparing inferred values to the known values from the simulations reveals that the methods used to determine the fitness effect and establishment times are accurate.

12.1 Simulation details

Parameters. The parameters used for the simulation are as follows

Parameter	Symbol	Value
Bottleneck Population size	N_b	5×10^7
Saturation Population size	N_s	$\sim 1.3 \times 10^{10}$
Effective population size	N	4×10^8
Generations between "transfers"	T	8
Total beneficial mutation rate	U_b	10^{-5}
Distribution of fitness effects	$\rho(s)$	`uniform(0, 0.125)`
Generations of prior growth	—	48
Generations of evolution	—	120
Number of barcode lineages	—	$500,000$
"Reads" per time point	R	3×10^7
Efficiency of PCR	β	0.02

Growth and bottlenecks. Growth is performed in discrete generations where the population of cells in the $t+1$ generation follows from that in the t generation via

$$\texttt{Number cells at } t+1 \texttt{ from single cell at } t = X(2) \qquad \text{If neutral} \tag{108}$$

$$\texttt{Number cells at } t+1 \texttt{ from single cell at } t = X(2(1+s)) \qquad \text{If has mutation with } s \tag{109}$$

After 8 such doublings the population is at saturation where the number of cells in our given barcode is n_s. Those cells are then Poisson sampled at a rate $\beta = 0.02$ to produce the k_b copies of "template DNA" that are present for amplification. These k copies undergo 23 rounds of doubling (modeling PCR) using the same procedure as before, namely $k_{t+1} = X(2k_t)$. The number of copies of each barcode at the end of this amplification stage is k_s. (Note that the details of the number of "cycles" is not very relevant: all noise comes from the early rounds when numbers are small hence things are left unaffected whether we have 10 or 23 cycles). The number k_s of copies of template DNA after "amplification" physically corresponds to the number of copies of the short product DNA that would be sent to the sequencer.

The next 8-generation cycle is started by sampling a fraction $\Delta = 2^{-8}$ of the saturated population and again allowing the cells to grown up by a factor of $\sim 2^8$ and once again performing the simulated versions of PCR and sequencing.

Modelling sequencing. Reads for each barcode r are generated by further Poisson sampling each barcode at a rate determined by its frequency in the short product DNA and the read depth R namely

$$r = X\left(R \times \frac{k_s}{\sum_i k_s(i)}\right) \tag{110}$$

where i indexes all barcodes. The combined levels of noise from the simulation performed this way is the same as measured for the experimental data.

Mutation. Each generation a cell has a probability $\mu(s)ds = U_b ds$ of getting a mutation in range $s, s+ds$. This is achieved via a Poisson sample with mean $U_b = 10^{-5}$. If the sampled yields a positive result (if it does, it is likely to be a singleton) this cell has its fitness effect drawn from a uniform distribution in the range $[0, 0.125]$. To model exclusively the effect of single mutants, once a cell gets one beneficial mutation it cannot mutate again. The distribution of mutation rates across fitnesses was chosen to be uniform because this offers the best chance of observing where biases due to detectability enter. The upper limit $s = 0.125$ was chosen since this is roughly the scale of the largest mutations observed in our experiment.

Prior growth and pre-existing mutations. Insertion of each of the 500,000 barcodes is modeled by drawing a number of cells $c = X(10)$ where $X(n)$ is a Poisson sample with mean n. To include the effects of the the $T \sim 33$ generations of growth prior to barcoding each cell has a probability of having a beneficial mutation in this drawing process. The probability is $\epsilon \sim 600/(5 \times 10^6)$, which is the number of beneficial cells expected from $T = 33$ generations of growth from a single cell up to 10^{10} cells with a beneficial mutation rate of 10^{-6} assuming they remain neutral over this growth process. These 5×10^6 cells labelled with 500,000 barcodes are then grown for $T = 16$ generations and permitted to undergo mutation. This common pool is then sampled into two independent simulations which are "evolved" for 120 generations.

12.2 Simulation results

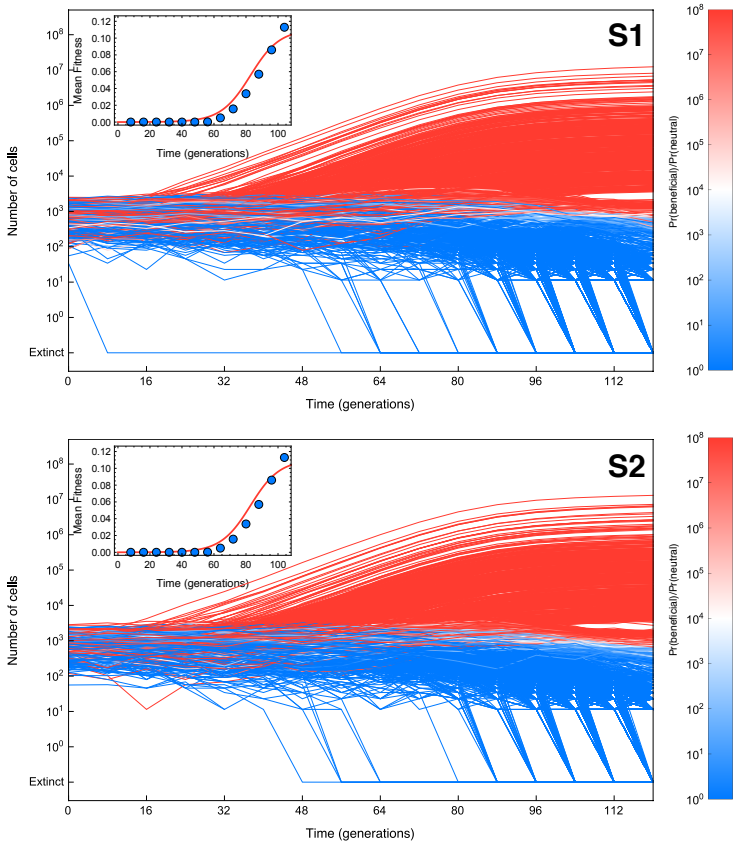


Figure 44: The trajectories of a subset of the 500,000 lineages colored according to to their probability of harboring a beneficial mutation for the two replicate simulations S1 and S2. In S1 7,621 adaptive lineages are identified by $t = 120$ and in S2, 7,747. Inset: The mean fitness inferred from the decline of neutral lineages (blue circles) agrees with the mean fitness calculated from the beneficial lineages (red line). Note there appears to be an offset of ~ 4 generations which arises because the mean fitness inferred from decline of neutral lineages is measured between two time points t and $t + 1$ which is 8 generations later. It is plotted at the time point corresponding to $t + 1$ though more probably reflects the mean fitness somewhere in the previous 8 generations.

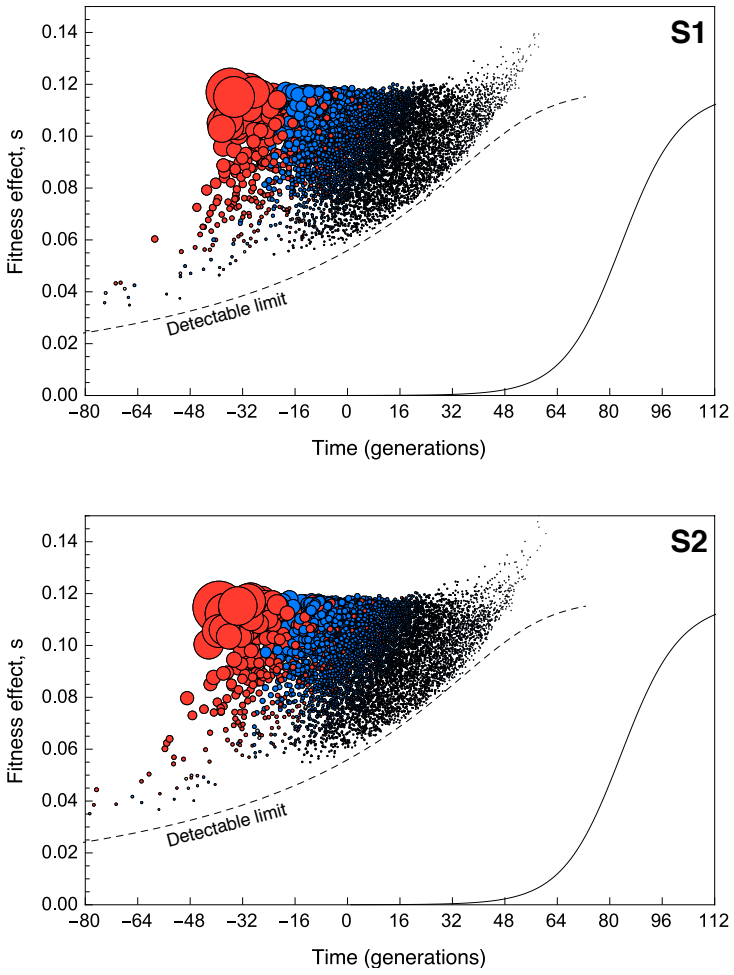


Figure 45: Joint distributions of the fitness effect (y-axis) and establishment times (x-axis) for the $\sim 7,000$ mutations identified as adaptive in each of the two replicate simulations. The number of pre-existing mutations that are sampled into both replicates is is 895 of which 620 establish and are identified as adaptive in both (red circles). Blue circles indicate lineages that we not identified as adaptive in both replicates. The area of the circles indicate the frequency of the lineage at $t = 88$. The detectable limit calculated using the result from Section 9.2 capture the limits of detection well.

13 Mathematical background

We now present the mathematical background needed to model the early dynamics of mutations, drift, and selection. While rare, the fate of a single cell and its descendants is independent of the others: this independence enables the use of general branching process methods. The overall goal is to be able determine the probability distribution over offspring numbers given various initial (and other) conditions.

13.1 Birth-death process

While the methods and the results we require to model lineages are more general (as we shall discuss) the simplest model to understand the statistics of mutations subject to drift and selection is a birth-death process for which in each unit of time there is a probability that a cell either divides or dies determined by birth and death rates.

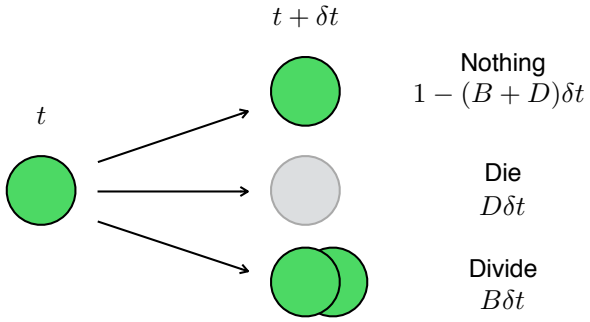


Figure 46: The birth-death branching process. In each time interval δt a cell can do one of three things (i) nothing, (ii) die and (iii) divide. For each cell these probabilities are independent.

At time t there are $n(t)$ cells. In a small interval of time δt, a cell can (i) do nothing, (ii) die or (iii) divide. The basic simplification is that cells are independent of one another (how many offspring one has does not influence the others): we can then write down how the number of cells changes in δt:

$$n(t + \delta t) = \sum_{j=1}^{n(t)} 1 + X_j \left(B\delta t\right) - X_j(D\delta t) \tag{111}$$

Where X_j are the random numbers of births / deaths for the jth cell and are drawn from (say) a Poisson distribution (though other distributions could equally be used), B the birth rate, D the death rate.

It is very useful to consider the dynamics of the moment generating function, defined as $\langle \exp(-\phi n) \rangle$. These dynamics can be obtained by propagating the expression for the moment generating function in time δt:

$$M(\phi(t + \delta t)) = \left\langle \exp\left[-\phi(t + \delta t)n(t + \delta t)\right] \right\rangle_{n(t+\delta t)} \tag{112}$$

Substituting in for $n(t + \delta t)$ in terms of $n(t)$ using the stochastic equation above, conditioning on having $n(t)$ and averaging over the stochastic variables X that occur between t and $t + \delta t$:

$$\left\langle \left\langle \exp\left[-\phi(t + \delta t)\left(\sum_{j=1}^{n(t)} 1 + X_j(B\delta t) - X_j(D\delta t)\right)\right] \right\rangle_X \right\rangle_{n(t)} \tag{113}$$

Since all the processes are independent of one another the averages over X can be written as products of the following form, where each one gives the generating function for a Poisson distribution

$$\langle \exp\left[\pm\phi(t+\delta t)X(B\delta t)\right]\rangle_X^{n(t)} = \exp\left[B\delta t\left(e^{\pm\phi(t+\delta t)}-1\right)\right] \tag{114}$$

By imposing that the form of the moment generating function remains invariant we impose a condition relating $\phi(t)$ to $\phi(t+\delta t)$ and therefore we realize that all the terms multiplying the n must be $-\phi(t)$.

We can now expand the terms in ϕ as a power series, and simplify to the case of interest for which the fitness, $s = B - D$, is the average difference between births and deaths in a single generation, is a small parameter. We then get:

$$-\phi_\alpha(t) = \phi(t+\delta t) - \delta t(B-D)\phi(t+\delta t) + \delta t\frac{B+D}{2}\phi^2(t+\delta t) \tag{115}$$

Which gives a differential equation for ϕ running backwards in time, or forwards in $\tau = T - t$

$$\boxed{\frac{\partial\phi}{\partial\tau} = s\phi - c\phi^2} \tag{116}$$

with $\phi(\tau = 0) = \phi(T)$. At this point we have generalized somewhat and replaced the variance per generation, which is $B+D$, for the simple model, by a more general variance in the number of offspring per average generation time, defining the parameter c to be half the variance in offspring number per generation: this depends on the specific birth death model — for example the fluctuations in the growth-dilution cycle of the experiments. (For continuous time division, $c = 0.5$). We now need to solve this backwards time equation to find $\phi(\tau)$ in terms of its initial condition $\phi(T)$. (Note that this is analogous to solving Komogorov's backwards time equation by Laplace transforming and using method of characteristics in the transformed variables, ϕ). Setting $\tau = T$ we obtain $\phi(0)$ in terms of $\phi(T)$. We then have the generating function at all times because we know initially there were $n(0)$ cells at time $t = 0$ and so initially the generating function was

$$M(\phi) = e^{-\phi(t=0)n(0)} \tag{117}$$

Substituting in for $\phi(0)$ in terms of $\phi(T)$ we obtain the generating function for all times. Note that by expanding in ϕ we have effectively ignored the discreteness of the number of cells. This gives correct results for small s and all n, including $n = 0$, except for $n = \mathcal{O}(\infty)$ for which details of the birth-death process matter.

13.2 Distribution of offspring from a single founding cell

It is instructive to consider the solution to this generating function in a few special cases the most important being that of a single mutant cell and its descendants. The solution to Eqn. 116, $\partial\phi/\partial\tau = s\phi - c\phi^2$, is

$$\phi(0) = \frac{a\phi(t)}{1+b\phi(t)} \tag{118}$$

where

$$a = e^{st} \quad \text{and} \quad b \sim (c/s)(e^{st}-1). \tag{119}$$

The generating function for the number of offspring at time is then simply

$$M(\phi) = \exp\left(-\frac{a\phi}{b\phi+1}\right) \tag{120}$$

where $\phi = \phi(T)$.

The extinction probability — simple from generating functions — follows immediately since $P(n=0) = M(\phi=\infty)$. This yields the extinction probability

$$P(n(t)=0) = e^{-a(t)/b(t)} \approx e^{-s/c}. \tag{121}$$

for long times: $st \gg 1$. For small s, which we are interested in, the exponential can be expanded and the establishment probability,

$$P_{est} = 1 - P_{extinct} \approx \frac{s}{c}. \tag{122}$$

We can also obtain the full distribution of offspring by inverting the generating function using the inverse Laplace transform. But as the one initial cell is small compared with the characteristic number for establishment, c/s, and we are interested in the behavior for numbers of this order or larger, the exponent will be small and the exponential can be expanded so the moment generating function reduces to

$$M(\phi) \approx 1 - \frac{a\phi}{b\phi + 1} \tag{123}$$

which is simply the Laplace transform of

$$P(n) = \left(1 - \frac{a}{b}\right)\delta(n) + \frac{a}{b^2}e^{-n/b}. \tag{124}$$

This is thus one way of deriving the expression in Eqn. 3. and showing that it is valid generally.

13.3 Distribution of offspring from n founding cells cell

The picture is similar starting with n_0 founding cells. The generating function is of the same form

$$M(\phi) = = \exp\left(-\frac{a\phi}{b\phi + 1}\right) \tag{125}$$

only now $a = n_0 e^{st}$. The extinction probability is then $P(n=0) \approx \exp(-n_0 s/c)$ which is why the establishment size (above which it is unlikely the mutation will fluctuate to extinction) is $n \sim c/s$. The inverse transform can be found exactly, giving a solution in terms of Bessel functions of the first kind (I_1):

$$P(n) = I_1\left(\frac{2\sqrt{an}}{b}\right)\exp\left(-\frac{n+a}{b}\right)\frac{1}{b}\sqrt{\frac{a}{n}} \tag{126}$$

however it is more instructive to invert it approximately: for large n using a saddle point approximation yields the probability distribution:

$$P(n) \approx \sqrt{\frac{a^{1/2}}{4\pi b n^{3/2}}}\exp\left[-\frac{(\sqrt{n}-\sqrt{a})^2}{b}\right] \tag{127}$$

In the limit of large n this develops an exponential rather than Gaussian tail with a characteristic decay length of $(c/s)(\exp(st)-1)$. Close to the mean however, it remains Gaussian with variance $\sim e^{2st}$. This expression is remarkable accurate over almost the entire range of n (excluding very small n, where it breaks down, see Figure 47).

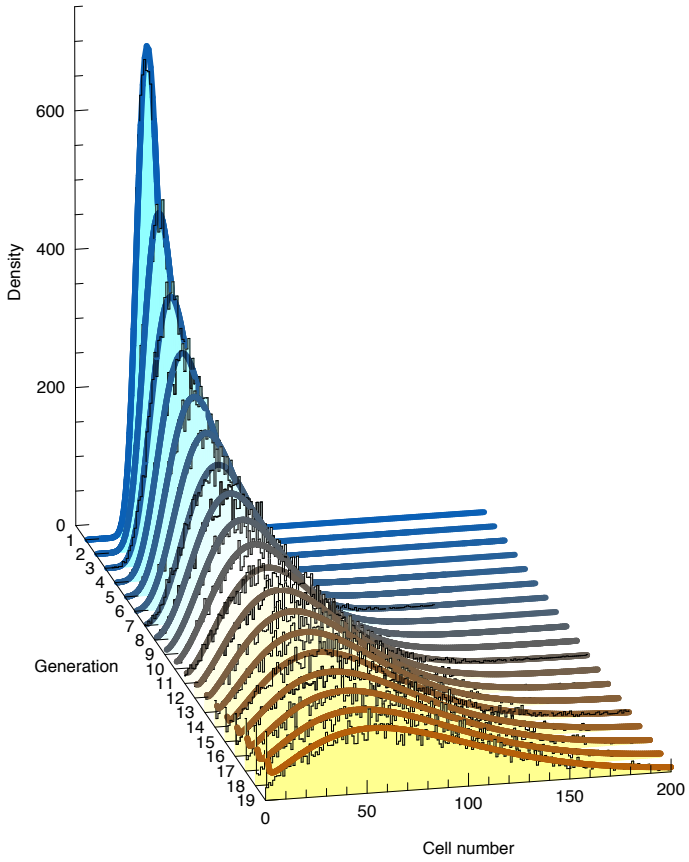


Figure 47: The distribution of lineage size under a birth death process. Initally $N = 30$ cells. Histograms are the result of 10,000 simulations whereby each cell can give rise to a Poisson distributed number of offspring in the next generation with mean $1 + s$ (here $s = 0.05$) and a variance around this of $2c = 1$. The analytic expressions obtained via the saddle-point approximation are plotted as the colored curves. Color is only used to help distinguish between time points.

13.4 The distribution of a mutant class being constantly fed by mutation

We next consider an approximately constant sized ancestral population that produces beneficial mutations at a constant rate $R = NU$, with each mutant behaving as those we just discussed. The distribution of offspring is now a convolution of all the mutants that enter over time. Since convolutions are products of the moment generating function, and since ϕ is the exponent, this translates to an integral of the original expression for $\phi = a\phi(t)/(1 + b\phi(t))$ over time:

$$\phi(0) = R \int_0^t \frac{a(t)\phi}{1 + b(t)\phi} dt = R \ln(\tilde{n}\phi + 1) \tag{128}$$

where $\tilde{n} = (c/s)[\exp(sT) - 1]$. It is more convenient to rescale n by $\tilde{n}$ and consider the generating function for $\nu = n/\tilde{n}$, which we denote $\langle \exp(-\eta\nu) \rangle$ yielding

$$M(\eta) = (1 + \eta)^R \tag{129}$$

This can be inverted exactly to yield the probability distribution over $\nu = n/\tilde{n}$

$$\rho(\nu)d\nu = \frac{d\nu}{\Gamma(R)} \frac{e^{-\nu}}{\nu^{1+R}} \tag{130}$$

Finally, we can derive the distribution distribution of establishment times for the constant feeding process by using the definition of the establishment time $\nu = e^{-s\tau}$. Substituting this in gives

$$\rho(\tau)d\tau = \frac{sd\tau}{\Gamma(R)} \exp\left(-Rs\tau - e^{-s\tau}\right). \tag{131}$$

which is the same as quoted in Eqn. 7

13.5 The distribution of a mutant class being exponentially fed by mutation

Next, we consider generation of second beneficial mutants. Instead of a constant population feeding mutants, we thus consider a population of cells growing exponentially

$$n_1 = Ne^{s_1 t} \tag{132}$$

that feeds beneficial mutations with fitness effects of s_2 at a rate of U per cell per generation. We want to determine the statistics of the sizes (number of cells) of the beneficial double-mutant sub-populations that enter and are destined to survive (Figure 13.5). The distribution of the total size of the resulting double mutant population can be obtained in an equivalent way to the constant feeding population case where we noticed that it is the convolution over time and hence an integral over ϕ:

$$\phi(0) = N\int_0^t e^{s_1 t}\frac{a(t)\phi}{b(t)\phi + 1}dt \tag{133}$$

where we note the additional factor of $e^{s_1 t}$ in the numerator and that now $a = e^{s_2 t}$ and $b = (c/s_2)(e^{s_2 t} - 1)$. This can be examined asymptotically, though involves some subtleties. It is more instructive to consider the statistics of single mutations and their occurrence times. Consider the first mutation (a) that enters at t_a and is destined to survive drift. It will reach a size

$$n_a = \frac{e^{s_2(t-t_a-\tau)}}{s_2} \tag{134}$$

after time t, where τ is a random variate from the establishment-time distribution of a single mutation (See Eqn.5) and is typically zero with an error of $\pm 1/s$ generations. What is the distribution of t_a? The cumulative number of divisions the feeding population undergoes between time t_1 and t_2 is

$$J = \frac{N(e^{s_1 t_2} - e^{s_1 t_1})}{s_1}. \tag{135}$$

For the first double mutant to occur in the interval dt_a around t_a requires that all previous mutations failed to establish, which has probability

$$(1 - Us_2)^J \approx \exp\left(-(NU/\alpha)e^{s_1 t_a}\right) \tag{136}$$

where $\alpha = s_1/s_2$. And the mutation that *is* destined to establish must occur in the interval dt_a which has probability

$$n_1 \times Us_2 \times dt_a = (NUs_2)\exp(s_1 t_a)\, dt_a \tag{137}$$

Thus the probability of the first mutation that is destined to establish occurring in the interval dt_a around t_a is

$$\rho(t_a)dt_a = (NUs_2)\exp\left(s_1 t_a - (NU/\alpha)e^{s_1 t_a}\right) dt_a \qquad (\text{for } t_a > 0) \tag{138}$$

This is plotted in Figure 49. This can also be cast as a distribution over the size of the first mutant $\rho(n_a)dn_a$

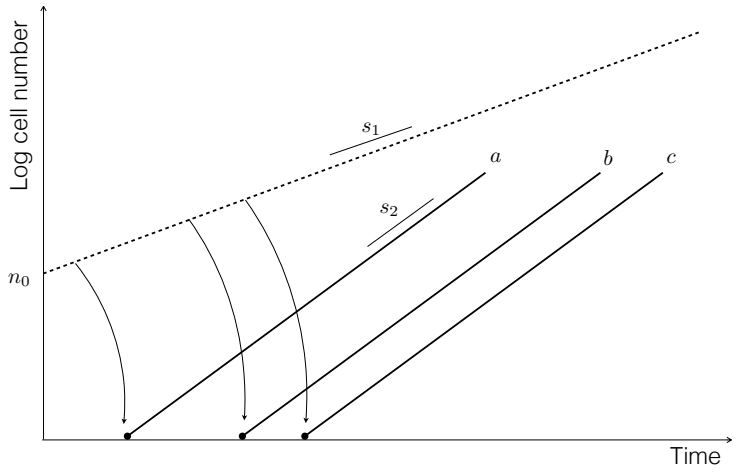


Figure 48: A population of cells growing like $ne^{s_1 t}$ feeds mutants at a rate of U_b, with fitness effects of s_2. Independent mutation enter and establish $(a, b, c...)$ from this growing feeding population. The relative sizes of the mutants is typically $1, 1/2, 1/3...$ if $s_1/s_2 \approx 1$.

by substituting using Eqn. 134 giving

$$\rho(n_a)dn_a = (NU)\tilde{n}^\alpha \exp\left(-\frac{NU}{\alpha}\left(\frac{n}{\tilde{n}}\right)^\alpha\right)\frac{dn_a}{n_a^{1+\alpha}} \tag{139}$$

where $\tilde{n} = e^{s_2 t}/s_2$.

One can think of the process of getting the second mutation, b, as the same as the first except time starts now at t_a and the initial population size is $n_0 e^{s_1 t_a}$. Hence the distribution of $\delta t_b = t_b - t_a$ is the same as for t_a with the replacement $n_0 \to n_0 e^{s_1 t_a}$ and so on for extensions to mutation c (Figure 49). What does this imply for the typical size of the mutants? If we take the median time t_k for all mutations, using Eqn. 134 we have that the size of the kth mutation to enter is

$$n_k \sim \left(\frac{e^{s_2(t-\tau)}}{s_2}\right)\left(\frac{NU}{\alpha}\right)^{1/\alpha}\frac{1}{k^{1/\alpha}} \tag{140}$$

with $k \in \left[1, (NU/\alpha)e^{s_1 t}\right]$

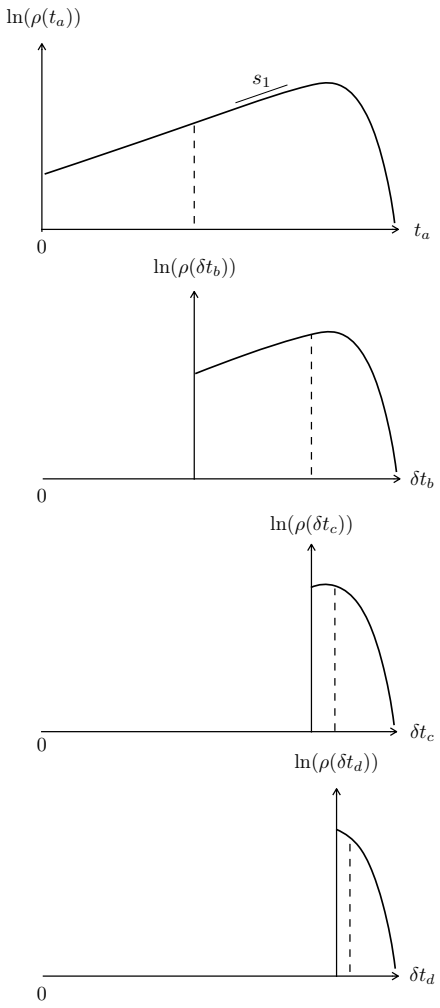


Figure 49: The establishment times for the first mutation (top) second mutation (second from top) and so on for mutations fed from an exponentially growing population.

References

[1] A L Goldstein and J H McCusker. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. Yeast, 15(14):1541–1553, October 1999.

[2] Henrik Albert, Emily C Dale, Elsa Lee, and David W Ow. Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. The Plant Journal, 7(4):649–659, 1995.

[3] Z Zhang and B Lutz. Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. Nucleic acids research, 30(17):e90, 2002.

[4] K C Kao and G Sherlock. Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. Nature Genetics, 40(12):1499–1504, December 2008.

[5] C B Brachmann, A Davies, G J Cost, E Caputo, J Li, P Hieter, and J D Boeke. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast, 14(2):115–132, January 1998.

[6] Kihoon Lee, Yu Zhang, and Sang Eun Lee. Saccharomyces cerevisiae ATM orthologue suppresses break-induced chromosome translocations. Nature, 454(7203):543–546, July 2008.

[7] C Verduyn, E Postma, W A Scheffers, and J P Van Dijken. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast, 8(7):501–517, July 1992.

[8] F M Ausubel, R Brent, R E Kingston, D D Moore, J G Seidman, J A Smith, and K Struhl. Current Protocols in Molecular Biology. Massachusetts General Hospital, Harvard Medical School. 1995.

[9] Lin Liu, Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law. Comparison of Next-Generation Sequencing Systems. BioMed Research International, 2012(7):1–11, July 2012.

[10] Juliane C Dohm, Claudio Lottaz, Tatiana Borodina, and Heinz Himmelbauer. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic acids research, 36(16):e105, September 2008.

[11] Gregory I Lang and Andrew W Murray. Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae. Genetics, 178(1):67–82, January 2008.

[12] J W Drake. A constant rate of spontaneous mutation in DNA-based microbes. Proc Natl Acad Sci U S A, 88(16):7160–7164, August 1991.

[13] Michael M Desai and Daniel S Fisher. Beneficial mutation selection balance and the effect of linkage on positive selection. Genetics, 176(3):1759–1798, July 2007.

[14] G I Lang, D Botstein, and M M Desai. Genetic variation and the fate of beneficial mutations in asexual populations. Genetics, 188(3):647–661, July 2011.

[15] Lília Perfeito, Lisete Fernandes, Catarina Mota, and Isabel Gordo. Adaptive mutations in bacteria: high rate and small effects. Science, 317(5839):813–815, August 2007.

[16] Michael M Desai, Daniel S Fisher, and Andrew W Murray. The speed of evolution and maintenance of variation in asexual populations. Curr Biol, 17(5):385–394, March 2007.

[17] Sarah B Joseph and David W Hall. Spontaneous mutations in diploid Saccharomyces cerevisiae: more beneficial than expected. Genetics, 168(4):1817–1825, December 2004.