

¹ Bayesian inference of relative fitness on high-throughput
² pooled competition assays

³ Manuel Razo-Mejia^{1,*}, Madhav Mani^{3,4}, and Dmitri Petrov^{1,2,5}

⁴ *¹Department of Biology, Stanford University*

⁵ *²Stanford Cancer Institute, Stanford University School of Medicine*

⁶ *³NSF-Simons Center for Quantitative Biology, Northwestern University*

⁷ *⁴Department of Molecular Biosciences, Northwestern University*

⁸ *⁵Chan Zuckerberg Biohub*

⁹ *Corresponding: *mrazo@stanford.edu*

¹⁰

¹¹ The tracking of lineage frequencies via DNA barcode sequencing enables the
¹² quantification of microbial fitness. However, experimental noise coming from
¹³ biotic and abiotic sources complicates the computation of a reliable inference.
¹⁴ We present a Bayesian pipeline to infer relative microbial fitness from high-
¹⁵ throughput lineage tracking assays. Our model accounts for multiple sources of
¹⁶ noise and propagates uncertainties throughout all parameters in a systematic way.
¹⁷ Furthermore, using modern variational inference methods based on automatic
¹⁸ differentiation, we are able to scale the inference to a large number of unique
¹⁹ barcodes. We extend this core model to analyze multi-environment assays,
²⁰ replicate experiments, and barcodes linked to genotypes. On simulations, our
²¹ method recovers known parameters within posterior credible intervals. This
²² work provides a generalizable Bayesian framework to analyze lineage tracking
²³ experiments. The accompanying open-source software library enables the adoption
²⁴ of principled statistical methods in experimental evolution.

²⁵ **Introduction**

²⁶ The advent of DNA barcoding—the ability to uniquely identify cell lineages with DNA
²⁷ sequences integrated at a specific locus—and high-throughput sequencing has opened new
²⁸ venues for understanding microbial evolutionary dynamics with an unprecedented level of
²⁹ temporal resolution^{1–3}. These experimental efforts rely on our ability to reliably infer the

30 relative fitness of an ensemble of diverse genotypes. Moreover, inferring these fitness values
31 over an ensemble of environmental conditions can help us determine the phenotypic diversity
32 of a rapid adaptation process⁴.

33 As with any other sequencing-based quantification, tracking lineages via DNA barcode
34 sequencing is inexorably accompanied by noise sources coming from experimental manipulation
35 of the microbial cultures, DNA extraction, and sequencing library preparation that involves
36 multiple rounds of PCR amplification, and the sequencing process itself. Thus, accounting for
37 the uncertainty when inferring the relevant parameters from the data is a crucial step to draw
38 reliable conclusions. Bayesian statistics presents a paradigm by which one can account for all
39 known sources of uncertainty in a principled way⁵. This, combined with the development
40 of modern Markov Chain Monte Carlo sampling algorithms⁶ and approximate variational
41 approaches⁷ have boosted a resurgence of Bayesian methods in different fields⁸.

42 We present a Bayesian inference pipeline to quantify the uncertainty about the parametric
43 information we can extract from high-throughput competitive fitness assays given a model
44 of the data generation process and experimental data. In these assays, the fitness of an
45 ensemble of genotypes is determined relative to a reference genotype^{3,4}. Figure 1(A) shows a
46 schematic of the experimental procedure in which an initial pool of barcoded strains are mixed
47 with a reference strain and inoculated into fresh media. After some time—usually, enough
48 time for the culture to saturate—an aliquot is transferred to fresh media, while the remaining
49 culture is used for DNA sequencing of the lineage barcodes. The time-series information of
50 the relative abundance of each lineage, i.e., the barcode frequency depicted in Figure 1(B), is
51 used to infer the relative fitness—the growth advantage on a per-cycle basis—for each lineage
52 with respect to the reference strain. The proposed statistical model accounts for multiple
53 sources of uncertainty when inferring the lineages' relative fitness values (see Section for
54 details on sources of uncertainty accounted for by the model). Furthermore, minor changes
55 to the core statistical model allow us to account for relevant experimental variations of these
56 competition assays. More specifically, in Section , we present a variation of the statistical
57 model to infer fitness on growth dilution cycles in multiple environments with proper error
58 propagation. Furthermore, as described in Section , our statistical model can account for
59 batch-to-batch differences when jointly analyzing multiple experimental replicates using a
60 Bayesian hierarchical model. Finally, a variant of these hierarchical models, presented in
61 Section , can account for variability within multiple barcodes mapping to equivalent genotypes
62 within the same experiment.

63 For all the model variations presented in this paper, we benchmark the ability of our pipeline
64 to infer relative fitness parameters against synthetic data generated from logistic growth
65 simulations with added random noise. A Julia package accompanies the present method to
66 readily implement the inference pipeline with state-of-the-art scientific computing software.

67 **Results**

68 **Experimental setup**

69 The present work is designed to analyze time-series data of relative abundance of multiple
70 microbial lineages uniquely identified by a DNA barcode^{3,4}. In these competition assays, an
71 ensemble of genotypes is pooled together with an unlabeled reference strain that, initially,
72 represents the vast majority ($\geq 90\%$) of the cells in the culture (see schematic in Figure 1(A)).
73 Furthermore, a fraction of labeled genotypes equivalent to the unlabeled reference strain—
74 hereafter defined as *neutral* lineages—are spiked in at a relatively high abundance ($\approx 3 - 5\%$).
75 The rest of the culture is left for the ensemble of genotypes of interest.

76 To determine the relative fitness of the ensemble of genotypes, a series of growth-dilution
77 cycles are performed on either a single or multiple environments. In other words, the cultures
78 are grown for some time; then, an aliquot is inoculated into fresh media for the next growth
79 cycle. This process is repeated for roughly 4-7 cycles, depending on the initial abundances of
80 the mutants and their relative growth rates. The DNA barcodes are sequenced at the end of
81 each growth cycle to quantify the relative abundance of each of the barcodes. We point the
82 reader to Kinsler et al. [4] for specific details on these assays for *S. cerevisiae* and to Ascensao
83 et al. [3] for equivalent assays for *E. coli*. Figure 1(B) presents a typical barcode trajectory
84 where the black trajectories represent the so-called *neutral* lineages, genetically equivalent
85 to the untagged ancestor strain that initially dominates the culture. These spiked-in neutral
86 lineages simplify the inference problem since the fitness metric of all relevant barcodes is
87 quantified with respect to these barcodes—thus referred to as *relative fitness*.

88 **Preliminaries on mathematical notation**

89 Before jumping directly into the Bayesian inference pipeline, let us establish the mathematical
90 notation used throughout this paper. We define (column) vectors as underlined lowercase
91 symbols such as

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}. \quad (1)$$

92 In the same way, we define matrices as double-underline uppercase symbols such as

$$\underline{\underline{A}} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{bmatrix}. \quad (2)$$

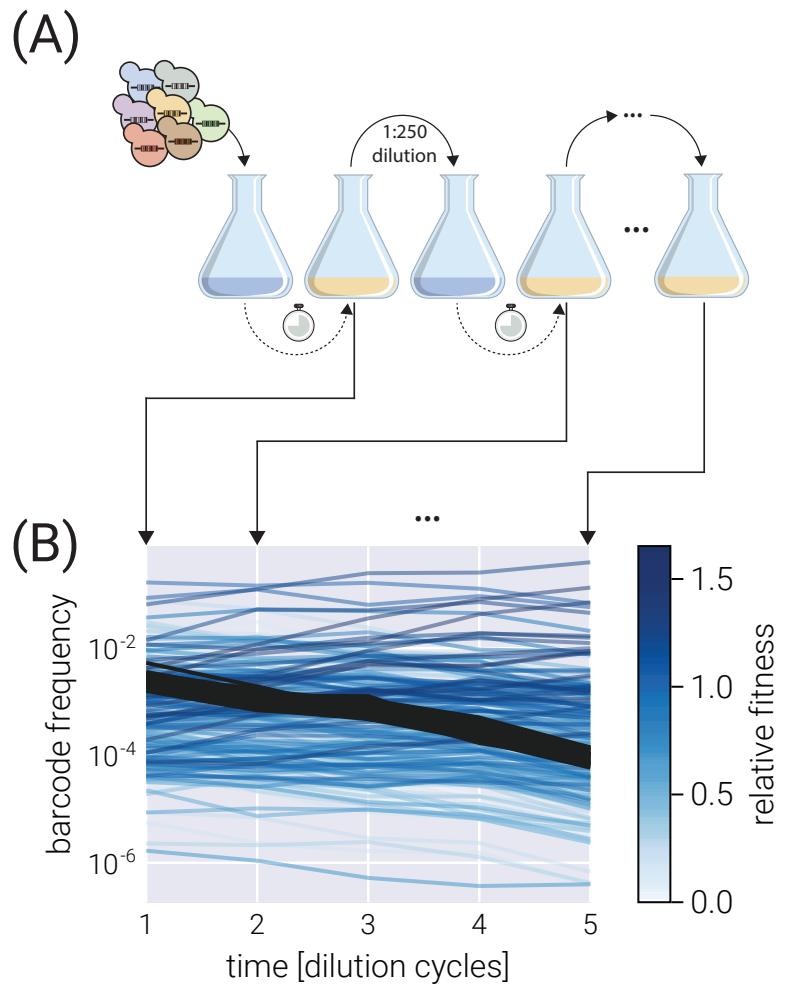


Figure 1. Typical competitive fitness experiment. (A) Schematic of the typical experimental design to determine the competitive fitness of an ensemble of barcoded genotypes. Genotypes are pooled together and grown over multiple growth-dilution cycles. At the end of each cycle, a sample is processed to generate a library for amplicon sequencing. (B) Typical barcode trajectory dataset. From each time point, the relative frequency of each barcode is determined from the total number of reads. Shades of blue represent different relative fitness. Darker gray lines define the typical trajectory of neutral lineages.

93 **Fitness model**

94 Empirically, each barcode frequency trajectory follows an exponential function of the
 95 form^{1,3,4}

$$f_{t+1}^{(b)} = f_t^{(b)} e^{(s^{(b)} - \bar{s}_t)\tau}, \quad (3)$$

96 where $f_t^{(b)}$ is the frequency of barcode b at the end of cycle number t , $s^{(b)}$ is the relative
 97 fitness with respect to the reference strain—the quantity we want to infer from the data— \bar{s}_t
 98 is the mean fitness of the culture at the end of cycle number t , and τ is the time pass between
 99 cycle t and $t + 1$. We can rewrite Equation 3 as

$$\frac{1}{\tau} \ln \frac{f_{t+1}^{(b)}}{f_t^{(b)}} = (s^{(b)} - \bar{s}_t). \quad (4)$$

100 Equation 4 separates the measurements—the barcode frequencies—from the unobserved
 101 (sometimes referred to as latent) parameters we want to infer from the data—the population
 102 mean fitness and the barcode relative fitness. This is ultimately the functional form used
 103 in our inference pipeline. Therefore, the relative fitness is computed by knowing the log
 104 frequency ratio of each barcode throughout the growth-dilution cycles.

105 The presence of the neutral lineages facilitates the determination of the population mean
 106 fitness value \bar{s}_t . Since every relative fitness is determined relative to the neutral lineage
 107 that dominates the culture, we define their fitness to be $s^{(n)} = 0$, where the superscript (n)
 108 specifies their neutrality. This means that Equation 4 for a neutral lineage takes the simpler
 109 form

$$\frac{1}{\tau} \ln \frac{f_{t+1}^{(n)}}{f_t^{(n)}} = -\bar{s}_t. \quad (5)$$

110 Therefore, we can use the data from these reference barcodes to directly infer the value of
 111 the population mean fitness.

112 It is important to notice that the frequencies $f_t^{(b)}$ are not the allele frequencies in the population
 113 (most of the culture is not sequenced since the reference strain is not barcoded), but rather
 114 the relative frequencies in the total number of sequencing reads. A way to conceptualize this
 115 subtle but important point is to assume exponential growth in the *number of cells* $N_t^{(b)}$ of
 116 the form

$$N_{t+1}^{(b)} = N_t^{(b)} e^{\lambda^{(b)}\tau}, \quad (6)$$

117 for every barcode b with growth rate $\lambda^{(b)}$. However, when we sequence barcodes, we do not
 118 directly measure the number of cells, but some number of reads $r_t^{(b)}$ that map to barcode b .
 119 In the simplest possible scenario, we assume

$$r_t^{(b)} \propto N_t^{(b)}, \quad (7)$$

120 where, importantly, the proportionality constant depends on the total number of reads for
 121 the library for cycle t , which might vary from library to library. Therefore, to compare the

number of reads between libraries at different time points, we must normalize the number of reads to the same scale. The simplest form is to define a relative abundance, i.e., a frequency with respect to the total number of reads,

$$f_t^{(b)} \equiv \frac{r_t^{(b)}}{\sum_{b'} r_t^{(b')}}. \quad (8)$$

This is the frequency Equation 3 describes.

Our ultimate objective is to infer the relative fitness $s^{(b)}$ for each of the M relevant barcodes in the experiment—hereafter referred to as $s^{(m)}$ to distinguish from the general $s^{(b)}$ and the neutral lineages $s^{(n)}$ relative fitness. To do so, we account for the three primary sources of uncertainty in our model:

1. Uncertainty in the determination of frequencies. Our model relates frequencies between adjacent growth-dilution cycles to the fitness of the corresponding strain. However, we do not directly measure frequencies. Instead, our data for each barcode consists of a length T vector of counts $\underline{r}^{(b)}$ for each of the T cycles in which the measurements were taken.
2. Uncertainty in the value of the population mean fitness. We define neutral lineages to have fitness $s^{(n)} = 0$, helping us anchor the value of the population mean fitness \bar{s}_t for each pair of adjacent growth cycles. Moreover, we take this parameter as an empirical parameter to be obtained from the data, meaning that we do not impose a functional form that relates \bar{s}_t to \bar{s}_{t+1} . Thus, we must infer the $T - 1$ values of this population mean fitness with their uncertainty that must be propagated to the value of the mutants' relative fitness.
3. Uncertainty in each of the mutants' fitness values.

To account for all these sources of uncertainty in a principled way, in the next section, we develop a Bayesian inference pipeline.

Bayesian inference

As defined in Section , our ultimate objective is to infer the vector of relative fitness values

$$\underline{s}^M = (s^{(1)}, s^{(2)}, \dots, s^{(M)})^\dagger, \quad (9)$$

where \dagger indicates the transpose. Our data consists of an $T \times B$ matrix $\underline{\underline{R}}$, where $B = M + N$ is the number of unique barcodes given by the sum of the number of unique, relevant barcodes we care about, M , and the number of unique neutral barcodes, N , and T is the number of growth cycles where measurements were taken. The data matrix is then of the form

$$\underline{\underline{R}} = \begin{bmatrix} - & r_1 & - \\ - & r_2 & - \\ \vdots & & \\ - & r_T & - \end{bmatrix}, \quad (10)$$

151 where each row \underline{r}_t is a B -dimensional array containing the raw barcode counts at cycle t . We
 152 can further split each vector \underline{r}_t into two vectors of the form

$$\underline{r}_t = \begin{bmatrix} \underline{r}_t^N \\ \underline{r}_t^M \end{bmatrix}, \quad (11)$$

153 i.e., the vector containing the neutral lineage barcode counts \underline{r}_t^N and the corresponding vector
 154 containing the mutant barcode counts \underline{r}_t^M . Following the same logic, matrix $\underline{\underline{R}}$ can be split
 155 into two matrices as

$$\underline{\underline{R}} = \begin{bmatrix} \underline{\underline{R}}^N & \underline{\underline{R}}^M \end{bmatrix}, \quad (12)$$

156 where $\underline{\underline{R}}^N$ is a $T \times N$ matrix with the barcode reads time series for each neutral lineage and
 157 $\underline{\underline{R}}^M$ is the equivalent $T \times M$ matrix for the non-neutral lineages.

158 Our objective is to compute the joint probability distribution for all relative fitness values
 159 given our data. We can express this joint posterior distribution using Bayes theorem as

$$\pi(\underline{s}^M | \underline{\underline{R}}) = \frac{\pi(\underline{\underline{R}} | \underline{s}^M)\pi(\underline{s}^M)}{\pi(\underline{\underline{R}})}, \quad (13)$$

160 where hereafter $\pi(\cdot)$ defines a probability density function. When defining our statistical
 161 model, we need not to focus on the denominator on the right-hand side of Equation 13. Thus,
 162 we can write

$$\pi(\underline{s}^M | \underline{\underline{R}}) \propto \pi(\underline{\underline{R}} | \underline{s}^M)\pi(\underline{s}^M). \quad (14)$$

163 However, when implementing the model computationally, the normalization constant on the
 164 right-hand side of Equation 13 must be computed. This can be done from the definition of
 165 the model via an integral of the form

$$\pi(\underline{\underline{R}}) = \int d^M \underline{s}^M \pi(\underline{\underline{R}} | \underline{s}^M)\pi(\underline{s}^M), \quad (15)$$

166 also known as a marginalization integral. Hereafter, differentials of the form d^n imply a
 167 n -dimensional integral.

168 Although Equation 13 and Equation 14 seem simple enough, recall that Equation 3 relates
 169 barcode frequency values and the population mean fitness to the mutant relative fitness.
 170 Therefore, we must include these nuisance parameters as part of our inference problem. We
 171 direct the reader to the supplementary materials for the exact definitions of these parameters.
 172 Here, it suffices to say that the inference problem must include the vector \bar{s}_T of all population
 173 mean fitness values and the matrix $\underline{\underline{F}}$ of all barcode frequencies within the sequencing data.
 174 With these nuisance variables in hand, the full inference problem we must solve takes the
 175 form

$$\pi(\underline{s}^M, \bar{s}_T, \underline{\underline{F}} | \underline{\underline{R}}) \propto \pi(\underline{\underline{R}} | \underline{s}^M, \bar{s}_T, \underline{\underline{F}})\pi(\underline{s}^M, \bar{s}_T, \underline{\underline{F}}). \quad (16)$$

176 To recover the marginal distribution over the non-neutral barcodes relative fitness values, we
 177 can numerically integrate out all nuisance parameters, i.e.,

$$\pi(\underline{s}^M | \underline{\underline{R}}) = \int d^{T-1} \bar{s}_T \int d^B f_1 \cdots \int d^B f_T \pi(\underline{s}^M, \bar{s}_T, \underline{\underline{F}} | \underline{\underline{R}}). \quad (17)$$

178 **Factorizing the posterior distribution**

179 The left-hand side of Equation 16 is extremely difficult to work with. However, we can take
 180 advantage of the structure of our inference problem to rewrite it in a more manageable form.
 181 Specifically, the statistical dependencies of our observations and latent variables allow us
 182 to factorize the joint distribution into the product of multiple conditional distributions. To
 183 gain some intuition about this factorization, let us focus on the inference of the population
 184 mean fitness values \bar{s}_T . Equation 5 relates the value of the population mean fitness to the
 185 neutral lineage frequencies and nothing else. This suggests that when writing the posterior
 186 for these population mean fitness parameters, we should be able to condition it only on the
 187 neutral lineage frequency values, i.e., $\pi(\bar{s}_T | \underline{F}^N)$. We point the reader to Section for the
 188 full mathematical details on this factorization. For our purpose here, it suffices to say we
 189 can rewrite the joint probability distribution as a product of conditional distributions of the
 190 form

$$\pi(\underline{s}^M, \bar{s}_T, \underline{F} | \underline{R}) = \pi(\underline{s}^M | \bar{s}_T, \underline{F}^M) \pi(\bar{s}_T | \underline{F}^N) \pi(\underline{F} | \underline{R}). \quad (18)$$

191 Written in this form, Equation 18 captures the three sources of uncertainty listed in Section in
 192 each term. Starting from right to left, the first term on the right-hand side of Equation 18
 193 accounts for the uncertainty when inferring the frequency values given the barcode reads.
 194 The second term accounts for the uncertainty in the values of the mean population fitness
 195 at different time points. The last term accounts for the uncertainty in the parameter we
 196 care about—the mutants' relative fitnesses. We refer the reader to Section for an extended
 197 description of the model with specific functional forms for each term on the left-hand side
 198 of Equation 18 as well as the extension of the model to account for multiple experimental
 199 replicates or hierarchical genotypes.

200 **Variational Inference**

201 One of the technical challenges to the adoption of Bayesian methods is the analytical in-
 202 tractability of integrals such as that of Equation 17. Furthermore, even though efficient
 203 Markov Chain Monte Carlo (MCMC) algorithms such as Hamiltonian Montecarlo can numeri-
 204 cally perform this integration⁶, the dimensionality of the problem in Equation 18 makes an
 205 MCMC-based approach prohibitively slow.

206 To overcome this computational limitation, we rely on the recent development of the automatic
 207 differentiation variational inference algorithm (ADVI)⁷. Briefly, when performing ADVI, our
 208 target posterior distribution $\pi(\theta | \underline{R})$, where $\theta = (\underline{s}^M, \bar{s}_T, \underline{F})$, is replaced by an approximate
 209 posterior distribution $q_\phi(\theta)$, where ϕ fully parametrizes the approximate distribution. As
 210 further explained in Section , the numerical integration problem is replaced by an optimization
 211 problem of the form

$$q_\phi^*(\theta) = \min_{\phi} D_{KL}(q_\phi(\theta) || \pi(\theta | \underline{R})), \quad (19)$$

212 where D_{KL} is the Kulback-Leibler divergence. In other words, the complicated high-
 213 dimensional numerical integration problem is transformed into a much simpler problem

214 of finding the value of the parameters ϕ such that Equation S4 is satisfied as best as possible
215 within some finite computation time. Although to compute Equation S4, we require the
216 posterior distribution we are trying to approximate $\pi(\theta | \underline{R})$, it can be shown that maximiz-
217 ing the so-called evidence lower bound (ELBO)⁹—equivalent to minimizing the variational
218 free energy¹⁰—is mathematically equivalent to performing the optimization prescribed by
219 Equation S4. We direct the reader to Section for a short primer on variational inference.

220 This work is accompanied by the Julia library BarBay.jl that makes use of the implementation
221 of both MCMC-based integration as well as ADVI optimization to numerically approximate
222 the solution of Equation 17 within the Julia ecosystem¹¹.

223 Inference on a single dataset

224 To assess the inference pipeline performance, we applied it to a simulated dataset with known
225 ground truth relative fitness values (See Section for details on simulation). Figure 2(A)
226 shows the structure of the synthetic dataset. The majority of barcodes of interest (faint
227 color lines) are adaptive compared to the neutral barcodes ($s^{(m)} > 0$). Although the barcode
228 frequency trajectories look relatively smooth, our fitness model requires the computation of
229 the log frequency ratio between adjacent time points as derived in Equation 4. Figure 2(B)
230 shows such data transformation where we can better appreciate the observational noise input
231 into our statistical model. This noise is evident for the darker lines representing the neutral
232 barcodes since all of these lineages are assumed to be identically distributed.

233 To visualize the performance of our inference pipeline in fitting our fitness model to the
234 observed data, we compute the so-called posterior predictive checks (PPC). In short, the
235 PPC consists of repeatedly generating synthetic datasets in agreement with the results from
236 the inference results. In other words, we use the resulting parameter values from the ADVI
237 inference to generate possible datasets in agreement with the inferred values (See Section for
238 further details on these computations). Figure 2(C) shows these results for all neutral lineages
239 (upper left corner plot) and a few representative non-neutral barcodes. The different color
240 shades represent the 95%, 68%, and 5% credible regions, i.e., the regions where we expect
241 to find the data with the corresponding probability—or in terms of our parameter, the $X\%$
242 credible region is the interval where we expect the true parameter value to lie with $X\%$
243 probability.

244 The main advantage of our method is this natural interpretability of these credible regions
245 where an $X\%$ credible region indeed captures the region of parameter space where we expect
246 with $X\%$ probability the actual value of the parameter lies given our statistical model, our
247 prior information, and the observed experimental data. A common mistake in the literature
248 is interpreting frequentist confidence intervals as Bayesian credible regions when they are
249 not equivalent¹². Frequentist confidence intervals and Bayesian credible regions are based
250 on fundamentally different philosophical approaches to statistics. Frequentist confidence
251 intervals represent the range of values that would contain the true population parameter with
252 a certain probability if the experiment was repeated many times. The confidence interval

253 does not represent the probability that the interval contains the true value. According to
 254 a specific model and prior information, Bayesian credible regions represent the range of
 255 values that contain the parameter with a certain posterior probability. The credible region
 256 directly represents the probability that the region contains the true value. So, frequentist
 257 confidence intervals cannot be interpreted as Bayesian credible regions because they have
 258 fundamentally different meanings. Treating an $X\%$ confidence interval like an $X\%$ credible
 259 region is fallacious since confidence intervals do not represent probabilistic coverage of the true
 260 value like credible regions. The intervals are generated through entirely different procedures.

261 To capture the global performance of the model, Figure 2(D) compares the known ground
 262 truth with the inferred relative fitness value for all barcodes of interest. There is an excellent
 263 degree of correspondence between these values, with the error bars representing the 68%
 264 credible region for the parameter value crossing the identity line for most barcodes. This
 265 latter point is made clear with Figure 2(E) where $\approx 90\%$ of ground truth fitness values fall
 266 within one standard deviation of the mean in the inferred posterior distributions.

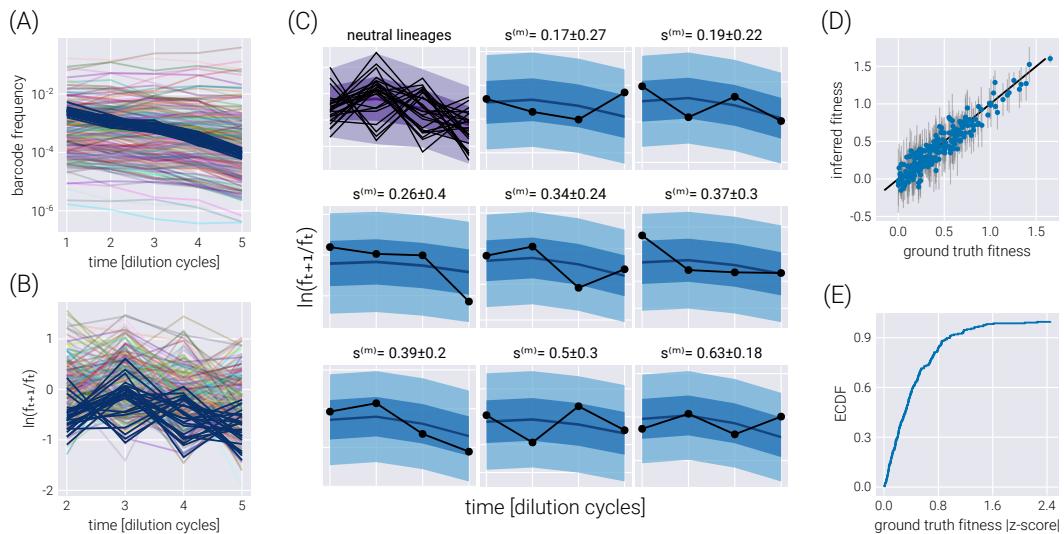


Figure 2. Single dataset inference. (A) Frequency trajectories that represent the raw data going into the inference. (B) Log frequency ratio between two adjacent time points used by the inference pipeline. Darker lines represent the neutral barcodes. These transformed data are much more noisy than the seemingly smooth frequency trajectories. (C) Examples of the posterior predictive checks for all neutral lineages (upper left panel) and a subset of representative mutant lineages. Shaded regions represent the 95%, 68%, and 5% credible regions for the data. The reported errors above the plot represent the 68% credible region on the mutant relative fitness marginal distribution. (D) Comparison between the ground truth fitness value from the logistic-growth simulation and the inferred fitness value. Gray error bars represent the 68% posterior credible region for the relative fitness values. (E) The empirical cumulative distribution function (ECDF) for the absolute z-score value of the ground truth parameter value within the inferred fitness posterior distribution.

267 **Fitness inference on multiple environments**

268 The fitness model in Equation 3 relates nuisance parameters such as the population mean
269 fitness and the barcode frequencies to the relative fitness parameter we want to infer from
270 the data. These dependencies imply that uncertainty on the estimates of these nuisance
271 parameters influences the inference of the relevant parameters. For example, imagine a
272 scenario where the neutral lineages data were incredibly noisy, leading to poor estimates of the
273 population mean fitness values \bar{s}_T . Since the relative fitness of any non-neutral barcode $s^{(m)}$ is
274 determined with respect to these neutral barcodes, not accounting for the lack of precision in
275 the value of the population mean fitness would result in misleading estimates of the accuracy
276 with which we determine the value of the parameter we care about. Thus, propagating these
277 sources of uncertainty in nuisance parameters is vital to generate an unbiased estimate of
278 the relevant information we want to extract from the data. One of the benefits of Bayesian
279 methods is the intrinsic error propagation embedded in the mathematical framework. For
280 our previous example, the uncertainty on the value of the population mean fitness values is
281 propagated to the relative fitness of a non-neutral barcode since we defined a joint posterior
282 distribution over all parameters as fully expressed in Equation 16.

283 This natural error propagation can help us with the experimental design schematized in
284 Figure 3(A). Here, rather than performing growth-dilution cycles in the same environment,
285 the cells are diluted into a different environment. Thus, the uncertainty on the fitness estimate
286 for the previous environment must be propagated to that of the next one. To validate the
287 extension of our statistical model to this scenario, Figure 3(B) shows the trajectory of the log
288 frequency ratios between adjacent time points. The different colored regions correspond to the
289 different environments. For this simulation, the growth rate of Environment 2 was set to be,
290 on average, half of the average growth rate in Environment 1. Equivalently, the growth rate
291 in Environment 3 was set to be, on average, twice the average growth rate in Environment
292 1. Figure 3(C-E) show the correspondence between the simulation ground truth and the
293 inferred fitness values, where the error bars represent the 68% credible region. Figure 3(F)
294 summarizes the performance of our inference pipeline by showing the empirical cumulative
295 distribution functions for the absolute value of the ground truth fitness value z-score within
296 the posterior distribution. This plot shows that, overall, $\approx 75\%$ of inferred mean values fall
297 within one standard deviation of the ground truth. For completeness, Figure 3(G) shows the
298 posterior predictive checks for a few example barcodes.

299 **Accounting for experimental replicates via hierarchical models**

300 Our inference pipeline can be extended to account for multiple experimental replicates via
301 Bayesian hierarchical models¹³. Briefly, when accounting for multiple repeated measurements
302 of the same phenomena, there are two extreme cases one can use to perform the data analysis:
303 On the one hand, we can treat each measurement as entirely independent, losing the power
304 to utilize multiple measurements when trying to learn a single parameter. This can negatively
305 impact the inference since, in principle, the value of our parameter of interest should not

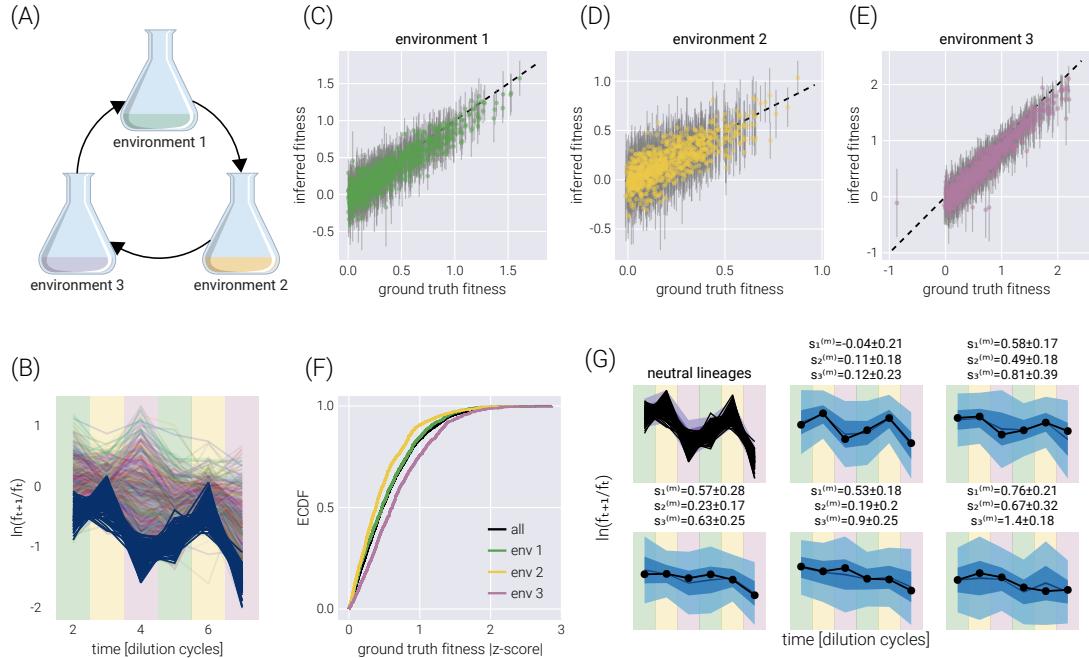


Figure 3. Multi-environment fitness inference. (A) Schematic of the simulated experimental design where growth-dilution cycles are performed into different environments for each cycle. (B) log frequency ratios between adjacent time points. Darker lines represent the neutral barcodes. The colors in the background demarcate the corresponding environment, matching colors in (A). Environment 2 is set to have, on average, half the growth rate of environment 1. Likewise, environment 3 is set to have, on average, twice the growth rate of environment 1. (C-E) Comparison between the ground truth fitness value from the logistic-growth simulation and the inferred fitness value for each environment. Gray error bars represent the 68% posterior credible region. (F) The empirical cumulative distribution function (ECDF) for the absolute z-score value of the ground truth parameter value within the inferred fitness posterior distribution for all fitness values (black line) and each environment individually (color lines). (G) Examples of the posterior predictive checks for all neutral lineages (upper left panel) and a subset of representative mutant lineages. Shaded regions surrounding the data represent the 95%, 68%, and 5% credible regions for the data. The reported errors above the plot represent the 68% credible region on the mutant relative fitness marginal distribution. Background colors match those of (A).

306 depend on the particular experimental replicate in question. However, this approach does
307 not allow us to properly “combine” the uncertainties in both experiments when performing
308 the inference. On the other hand, we can pool all data together and treat our different
309 experiments as a single measurement with higher coverage. This loses the subtle differences
310 due to biotic and abiotic batch effects, effectively halving the data that goes into our inference
311 problem.

312 Hierarchical models present a middle ground between these extremes. First, hierarchical
313 models rely on the definition of so-called *hyper-parameters*, that capture the parametric
314 inference we are interested in—for this inference problem, we have a hyper-fitness value
315 $\theta^{(m)}$ for each non-neutral barcode. Second, each experiment draws randomly from the
316 distribution of this hyper-parameter, allowing for subtle variability between experiments to be
317 accounted for—in the present inference pipeline, each experimental replicate gets assigned a
318 *local* fitness value $s_i^{(m)}$, where the extra sub-index indicates the i -th experimental replicate.
319 Conceptually, we can think of the local fitness for replicate i as being sampled from a
320 distribution that depends on the value of the global hyper-fitness value, i.e., $s_i^{(m)} \sim \pi_{\theta^{(m)}}$,
321 where the subindex $\theta^{(m)}$ indicates the distribution’s parametric dependence on the hyper-
322 fitness value. This way of interpreting the connection between the distribution $\pi_{\theta^{(m)}}$ and
323 the local fitness implies that a large replicate-to-replicate variability would lead to a broad
324 hyper-fitness distribution—implying a large uncertainty when determining the parameter that
325 characterizes the overall relative fitness. We point the reader to Section for the full definition
326 of the hierarchical model used in this section. Importantly, as schematized in Figure 4(A),
327 the influence between different experimental replicates runs both ways. First, the data from
328 one experimental replicate (R_k^M in the diagram) informs all local fitness values via the global
329 hyper-fitness (upper panel in Figure 4(A)). Second, the local fitness value is informed by the
330 data from all experimental replicates via the same global hyper-fitness parameter (lower panel
331 in Figure 4(A)).

332 To test the performance of this model, we simulated two experimental replicates with 1000
333 unique barcodes (see Figure 4(B-C)) where we randomly sampled a ground truth hyper-fitness
334 value $\theta^{(m)}$ for each barcode. We sampled a variation from this hyper-fitness value for each
335 experimental replicate $s_i^{(m)}$ to capture experimental batch effects. Figure 4(D) shows the
336 relationship between hyper-fitness and replicate fitness values for this simulation. The spread
337 around the identity line represents the expected batch-to-batch variation. The posterior
338 predictive checks examples in Figure 4(E) show that the hierarchical model can correctly fit
339 the data for each experimental replicate. Furthermore, Figure 4(F-G) show a high correlation
340 between the ground truth and the inferred fitness values. The empirical cumulative distribution
341 functions shown in Figure 4(H-I) reveal that for $\approx 75\%$ of the non-neutral barcodes, the
342 ground truth hyper-fitness values fall within one standard deviation from the mean value in
343 the posterior distributions.

344 As shown in Figure 5, the structure imposed by the hierarchical model schematized in
345 Figure 4(A), where we explicitly account for the connection between experimental replicates
346 can improve the quality of the inference. Inferred fitness values between experimental
347 replicates exhibit a stronger degree of correlation in the hierarchical model (Figure 5(A))

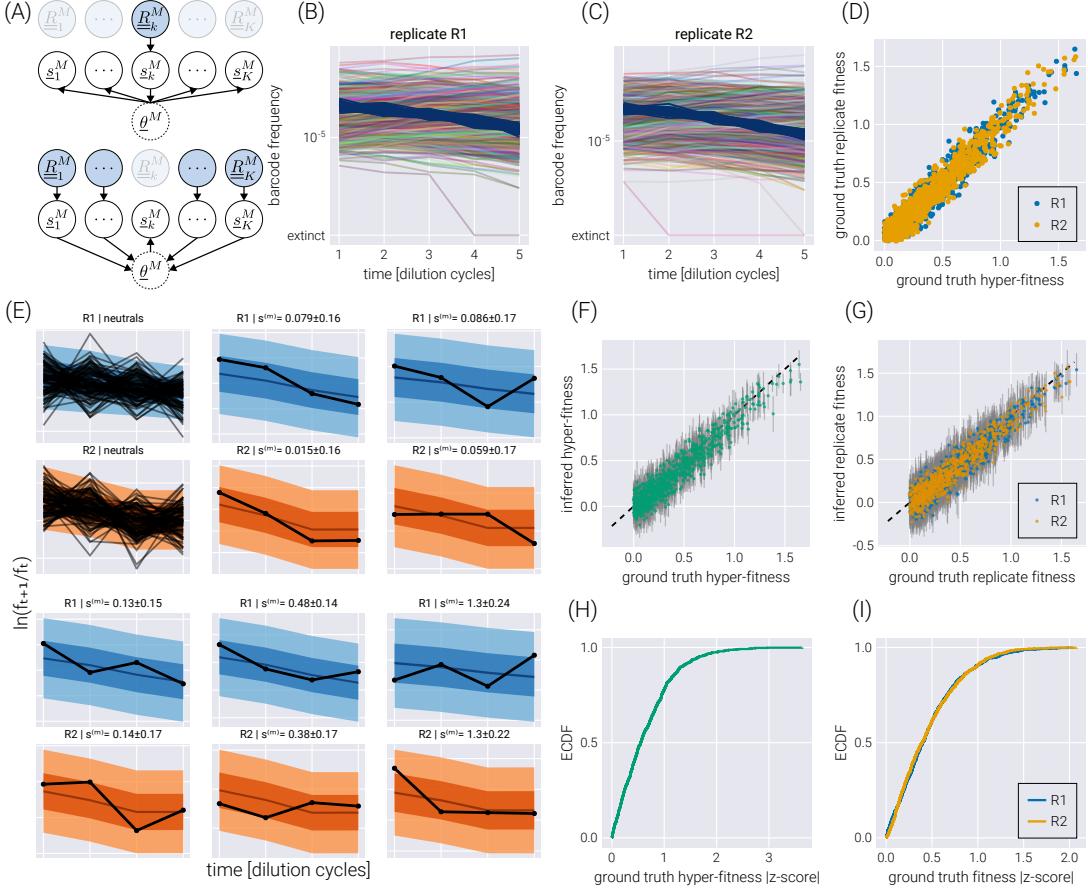


Figure 4. Hierarchical model on experimental replicates. (A) Schematic depiction of the interactions between local fitness values \underline{s}_k^M through the global hyper-fitness value θ^M for K hypothetical experimental replicates. The upper diagram shows how the data from replicate k informs all local fitness values via the hyper-fitness parameter. The lower panel shows the reverse, where all other datasets inform the local fitness value. (B-C) Simulated replicate datasets with 900 barcodes of interest and 100 neutral lineages. (D) Comparison between the simulation ground truth hyper-fitness and each replicate ground truth fitness. The scatter between parameters captures experimental batch effects. (E) Examples of the posterior predictive checks for all neutral lineages (upper left panels) and a subset of representative mutant lineages. Shaded regions from light to dark represent the 95%, 68%, and 5% credible regions. (F-G) Comparison between the simulation's ground truth hyper-fitness (F) and replicate fitness (G) values and the inferred parameters. Gray error bars represent the 68% posterior credible region. (H-I) The empirical cumulative distribution function (ECDF) for the absolute z-score value of the ground truth parameter value within the inferred hyper-fitness posterior distribution (H) and replicate fitness (I).

348 compared to conducting inference on each replicate independently (Figure 5(B)). Moreover,
349 when comparing the inferred hyper-fitness values—the objective parameter when performing
350 multiple experimental measurements—the hierarchical model outperforms averaging the
351 independent experimental replicates as shown in Figure 5(C) and (D).

352 **Accounting for multiple barcodes per genotype via hierarchical models**

353 Hierarchical models can also capture another experimental design in which multiple barcodes
354 map to the same or an equivalent genotype. As we will show, this many-to-one mapping
355 can improve the inference compared to the extreme cases of inferring the fitness of each
356 barcode independently or pooling the data of all barcodes mapping to a single genotype. As
357 schematized in Figure 6(A), a small modification of the base model allows us to map the
358 structure of our original model to that of a hierarchical model with a fitness hyperparameter
359 vector θ^G , where G is the number of genotypes in the dataset.

360 Figure 6(B) shows a single experimental replicate in which 90 genotypes were assigned
361 a random number of barcodes (a multinomial distribution with a mean of ten barcodes
362 per genotype) for a total of 900 non-neutral barcodes. To assess the performance of the
363 hierarchical model proposed in Figure 6(A), we performed inference using this hierarchical
364 model, as well as the two extreme cases of ignoring the connection between the barcodes
365 belonging to the same genotype—equivalent to performing inference using the model presented
366 in Figure 2(A) over the barcodes—or pooling the data of all barcodes belonging to the same
367 genotype into a single count—equivalent to performing inference using the model presented
368 in Figure 2(A) over the pooled barcodes. Figure 6(C-D) shows the comparison between
369 the simulation ground truth and the inferred values for these three cases. Not only do the
370 hierarchical model results show higher degrees of correlation with the ground truth, but the
371 error bars (representing the 68% credible regions) are smaller, meaning that the uncertainty
372 in the estimate of the parameter we care about decreases when using the hierarchical model.
373 The improvement in the prediction can be seen in Figure 6(F) where the empirical cumulative
374 distribution function of the absolute difference between the mean inferred value and the
375 simulation ground truth is shown for all three inference models. The hierarchical model's
376 curve ascends more rapidly, showing that, in general, the inferred values are closer to the
377 ground truth. For completeness, Figure 6(G) shows some examples of how the hierarchical
378 model can capture the raw log-frequency count observations.

379 **Discussion**

380 Experimental evolution of microbial systems has dramatically advanced our understanding
381 of the basic principles of biological evolution¹⁴. From questions related to the optimal
382 fine-tuning of gene expression programs¹⁵, to the dimensionality, geometry, and accessibility
383 of the adaptive fitness landscape explored by these rapidly adapting populations^{4,16}, to the
384 emergence of eco-evolutionary dynamics in a long-term evolution experiment¹⁷; for all of

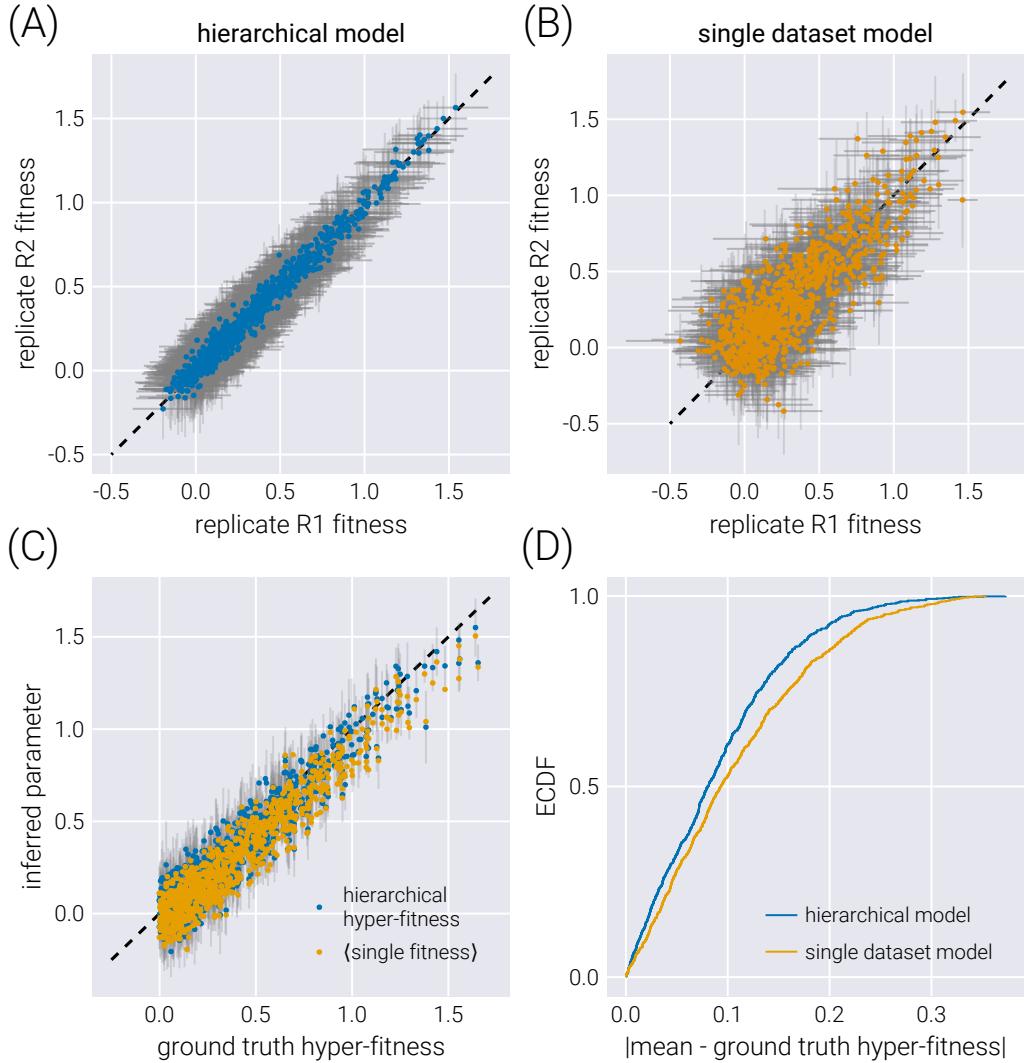


Figure 5. Comparison between hierarchical model and single dataset model. (A-B) comparison of inferred fitness values between experimental replicates when fitting a hierarchical model (A) or independently fitting each dataset (B). Gray error bars represent the 68% credible regions. (C) Comparison between the ground truth hyper-fitness value and the inferred parameters. The blue dots show the inferred hyper-fitness values when assuming a hierarchical model. Gray error bars show the 68% credible region for this inference. The yellow dots show the average of the mean inferred fitness values for the two experimental replicates. No error bars are shown for these, as it is inappropriate to compute one with two data points per non-neutral barcode. (D) Empirical cumulative distribution function (ECDF) of the absolute difference between the inferred mean and the ground truth hyper-fitness.

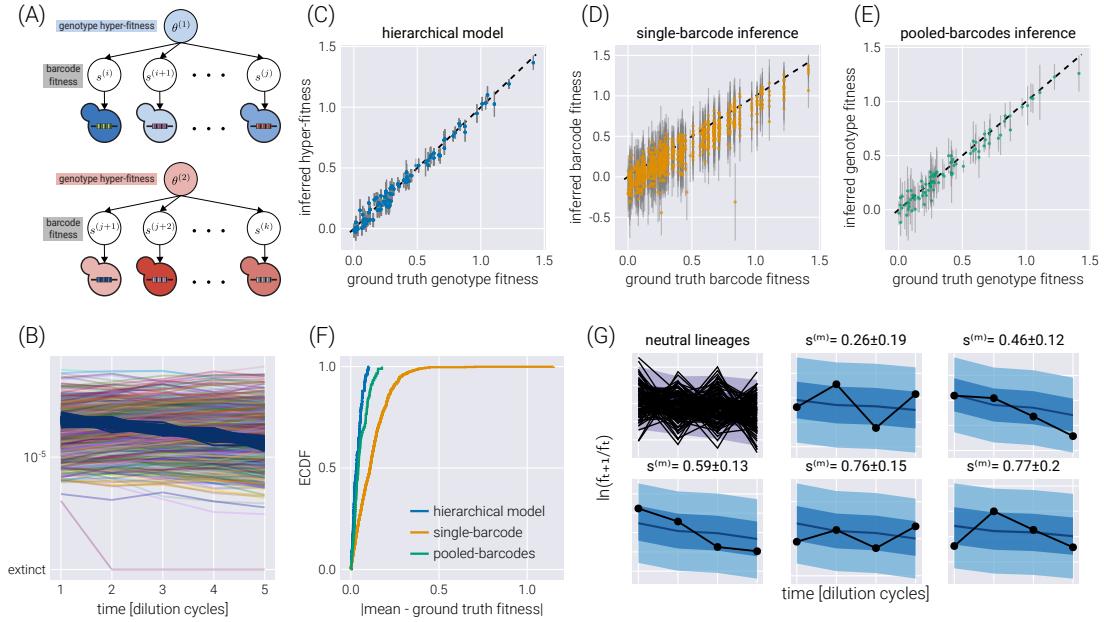


Figure 6. Hierarchical model for multiple barcodes per genotype. (A) Schematic depiction of the hierarchical structure for multiple barcodes mapping to a single genotype. A set of barcodes mapping to an equivalent genotype map to “local” fitness values $s^{(b)}$ that are connected via a hyper-fitness parameter for the genotype $\theta^{(g)}$. (B) Simulated dataset with 100 neutral lineages and 900 barcodes of interest distributed among 90 genotypes. (C-E) Comparison between the inferred and ground truth fitness values for a hierarchical model (C), a model where each barcode is inferred independently (D), and a model where barcodes mapping to the same genotype are pooled together (E). Gray error bars represent the 68% credible regions. (F) Empirical cumulative distribution function (ECDF) of the absolute difference between the inferred mean and the ground truth fitness values for all three models. (G) Examples of the posterior predictive checks for all neutral lineages (upper left panels) and a subset of representative mutant lineages. Shaded regions from light to dark represent the 95%, 68%, and 5% credible regions.

385 these and other cases, the microbial experimental platform combined with high-throughput
386 sequencing has been essential to tackling these questions with empirical data. This exciting
387 research area promises to improve as new culturing technologies¹⁸ as well as more complex
388 lineage barcoding schemes^{2,19}, are adopted.

389 For this data-heavy field, properly accounting for the uncertainty in parameters inferred from
390 experiments is vital to ensure the conclusions drawn are reliable. Bayesian statistics presents
391 a principled way to quantify this uncertainty systematically²⁰. Moreover, Bayesian analysis
392 offers a more natural way to interpret the role that probability theory plays when performing
393 data analysis compared to the often-misinterpreted frequentist methods²¹. Nevertheless, the
394 technical challenges associated with Bayesian analysis has limited its application. This is set
395 to change as recognition of the misuse of frequentist concepts such as the p-value is receiving
396 more attention²². Moreover, advances in numerical methods such as Hamiltonian Monte
397 Carlo⁶ and variational inference⁷ allows for complex Bayesian models to be fit to empirical
398 data.

399 In this paper, we present a computational pipeline to analyze lineage-tracking time-series data
400 for massive-parallel competition assays. More specifically, we fit a Bayesian model to infer
401 the fitness of multiple genotypes relative to a reference^{3,4}. The proposed model accounts for
402 multiple sources of uncertainty with proper error propagation intrinsic to Bayesian methods.
403 To scale the inference pipeline to large datasets with > 10,000 barcodes, we use the ADVI
404 algorithm⁷ to fit a variational posterior distribution. The main difference between our method
405 and previous inference pipelines, such as Li et al. [23], is that the present analysis provides
406 interpretable errors on the inferred fitness values. The reported uncertainty intervals—known
407 as credible regions—can be formally interpreted as capturing the corresponding probability
408 mass of finding the true value of the parameter given the model, the prior information, and
409 the data. Furthermore, minor modifications to the structure of the statistical model presented
410 in this work allow for the analysis of different experimental designs, such as growth-dilution
411 cycles in different environments, joint analysis of multiple experimental replicates of the same
412 experiment via hierarchical models, and a hierarchical model for multiple barcodes mapping
413 to equivalent genotypes. We validate our analysis pipeline on simulated datasets with known
414 ground truth, showing that the model fits the data adequately, capturing the ground truth
415 parameters within the posterior distribution.

416 It is important to highlight some of the consequences of the general experimental design and
417 the implicit assumptions within the proposed statistical model to analyze the resulting data.
418 First, the composition of the population is such that the initial fraction of the population
419 occupied by the barcoded genotypes is small—usually >90% of the initial population is
420 the non-labeled reference strain. This constraint is important as the fitness model used
421 to fit the time series data assumes that the tracked frequencies are $\ll 1$. Second, when
422 computing log frequency ratios, we can run into the issue of dividing by zero. This is a
423 common problem when dealing with molecular count data²⁴. Our model gets around this issue
424 by assuming that the frequency of any barcode cannot be, but still can get arbitrarily close
425 to, zero. Therefore, we implicitly assume that no lineage goes extinct during the experiment.
426 Moreover, the statistical model directly accounts for the uncertainty associated with having

427 zero barcode counts, increasing the corresponding uncertainty. Third, the models presented
428 in this paper require the existence of a labeled sub-population of barcoded reference strains.
429 These barcodes help determine the fitness baseline, as every fitness is quantified with respect
430 to this reference genotype. This experimental design constraint facilitates the inference of
431 the population mean fitness since most of the culture—the unlabeled reference genotype—is
432 not tracked. Finally, the presented statistical model assumes that relative fitness is solely a
433 constant of the environment and the genotype. Future directions of this work could extend
434 the fitness model to properly analyze data with time-varying or frequency-dependent fitness
435 values.

436 In total, the statistical model presented in this work and the software package accompanying
437 the paper allow for a principled way of quantifying the accuracy with which we can extract
438 relevant parametric information from large-scale multiplexed fitness competition assays. Fur-
439 thermore, the implementation of Bayesian models and their fitting via automatic differentiation
440 approaches opens the gate to extend this type of formal analysis to the data-rich literature in
441 experimental evolution and other high-throughput technologies applications.

442 Acknowledgements

443 We would like to thank Griffin Chure and Michael Betancourt for their helpful advice and
444 discussion. We would like to thank Karna Gowda, Spencer Farrell, and Shaili Mathur for
445 critical observations on the manuscript. This work was supported by the NIH/NIGMS,
446 Genomics of rapid adaptation in the lab and in the wild R35GM11816506 (MIRA grant),
447 the NIH, Unravelling mechanisms of tumor suppression in lung cancer R01CA23434903,
448 the NIH (PQ4), Quantitative and multiplexed analysis of gene function in cancer in vivo
449 R01CA23125303, the NIH, Genetic Determinants of Tumor Growth and Drug Sensitivity
450 in EGFR Mutant Lung Cancer R01CA263715, the NIH, Dissecting the interplay between
451 aging, genotype, and the microenvironment in lung cancer U01AG077922, the NIH, Genetic
452 dissection of oncogenic Kras signaling R01CA230025, and the CZ Biohub investigator program.
453 MRM was supported by the Schmidt Science Fellowship. MM was supported by The National
454 Science Foundation-Simons Center for Quantitative Biology at Northwestern University and
455 the Simons Foundation grant 597491. MM is a Simons Investigator. DP is a CZ Biohub
456 investigator.

457 References

- 458 ¹S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, and G. Sherlock,
459 “Quantitative evolutionary dynamics using high-resolution lineage tracking”, *Nature* **519**,
460 [181–186 \(2015\)](#).
- 461 ²A. N. Nguyen Ba, I. Cvijović, J. I. Rojas Echenique, K. R. Lawrence, A. Rego-Costa, X. Liu,
462 S. F. Levy, and M. M. Desai, “High-resolution lineage tracking reveals travelling wave of
463 adaptation in laboratory yeast”, *Nature* **575**, [494–499 \(2019\)](#).

- 464 ³J. A. Ascensao, K. M. Wetmore, B. H. Good, A. P. Arkin, and O. Hallatschek, "Quantifying
465 the local adaptive landscape of a nascent bacterial community", *Nature Communications*
466 **14**, 248 (2023).
- 467 ⁴G. Kinsler, K. Geiler-Samerotte, and D. A. Petrov, "Fitness variation across subtle environ-
468 mental perturbations reveals local modularity and global pleiotropy of adaptation", *eLife* **9**,
469 1–52 (2020).
- 470 ⁵S. R. Eddy, "What is Bayesian statistics?", *Nature Biotechnology* **22**, 1177–1178 (2004).
- 471 ⁶M. Betancourt, "A Conceptual Introduction to Hamiltonian Monte Carlo", ArXiv (2017).
- 472 ⁷A. Kucukelbir, D. Tran, R. Ranganath, A. Gelman, and D. M. Blei, *Automatic Differentiation
473 Variational Inference*, (Mar. 2, 2016) (visited on 07/07/2023), preprint.
- 474 ⁸B. Efron, "Bayes' Theorem in the 21st Century", *Science* **340**, 1177–1178 (2013).
- 475 ⁹D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, (May 1, 2014) <http://arxiv.org/abs/1312.6114> (visited on 11/21/2022), preprint.
- 477 ¹⁰S. Gottwald and D. A. Braun, "The two kinds of free energy and the Bayesian revolution",
478 *PLOS Computational Biology* **16**, e1008420 (2020).
- 479 ¹¹H. Ge, K. Xu, and Z. Ghahramani, "Turing: A Language for Flexible Probabilistic Inference",
480 in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and
481 Statistics* (Mar. 31, 2018), pp. 1682–1690.
- 482 ¹²R. D. Morey, R. Hoekstra, J. N. Rouder, M. D. Lee, and E.-J. Wagenmakers, "The fallacy
483 of placing confidence in confidence intervals", *Psychonomic Bulletin & Review* **23**, 103–123
484 (2016).
- 485 ¹³M. J. Betancourt and M. Girolami, *Hamiltonian Monte Carlo for Hierarchical Models*,
486 (Dec. 3, 2013) <http://arxiv.org/abs/1312.0906> (visited on 07/20/2023), preprint.
- 487 ¹⁴E. Kussell, "Evolution in Microbes", *Annu. Rev. Biophys* **42**, 493–514 (2013).
- 488 ¹⁵E. Dekel and U. Alon, "Optimality and evolutionary tuning of the expression level of a
489 protein", *Nature* **436**, 588–592 (2005).
- 490 ¹⁶T. Maeda, J. Iwasawa, H. Kotani, N. Sakata, M. Kawada, T. Horinouchi, A. Sakai, K.
491 Tanabe, and C. Furusawa, "High-throughput laboratory evolution reveals evolutionary
492 constraints in *Escherichia coli*", *Nature Communications* **11**, 5970 (2020).
- 493 ¹⁷B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai, "The dynamics
494 of molecular evolution over 60,000 generations", *Nature*, 10.1038/nature24287 (2017).
- 495 ¹⁸T. Jagdish and A. N. Nguyen Ba, "Microbial experimental evolution in a massively multi-
496 plexed and high-throughput era", *Current Opinion in Genetics & Development* **75**, 101943
497 (2022).
- 498 ¹⁹D. Yang et al., "Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor
499 evolution", *Cell* **185**, 1905–1923.e25 (2022).
- 500 ²⁰A. Gelman and C. R. Shalizi, "Philosophy and the practice of Bayesian statistics", *Statistics
501 math.ST*, 36 (2010).

- 502 ²¹J. VanderPlas, “Frequentism and Bayesianism: A Python-driven Primer”, [ArXiv](#), 1–9 (2014).
- 503 ²²R. Nuzzo, “Statistical errors”, [Nature](#) **506** (2014).
- 504 ²³F. Li, J. Tarkington, and G. Sherlock, “Fit-Seq2.0: An Improved Software for High-
505 Throughput Fitness Measurements Using Pooled Competition Assays”, [Journal of Molecular
506 Evolution](#), [10.1007/s00239-023-10098-0](#) (2023).
- 507 ²⁴D. R. Lovell, X.-Y. Chua, and A. McGrath, “Counts: an outstanding challenge for log-
508 ratio analysis of compositional data in the molecular biosciences”, [NAR Genomics and
509 Bioinformatics](#) **2**, lqaa040 (2020).

510 **Supplementary Materials**

511 **Table of contents**

512	Supplementary Materials	22
513	Primer on Variational Inference	22
514	ADVI algorithm	26
515	Defining the Bayesian model	27
516	Frequency uncertainty $\pi(\underline{F} \underline{R})$	27
517	Population mean fitness uncertainty $\pi(\bar{s}_T \underline{F}, \underline{R})$	29
518	Mutant relative fitness uncertainty $\pi(\underline{s}_T^M \bar{s}_T, \underline{F}, \underline{R})$	33
519	Hierarchical models for multiple experimental replicates	35
520	Defining prior probabilities	38
521	Posterior predictive checks	40
522	Logistic growth simulation	41

523 **Primer on Variational Inference**

524 In this section, we will briefly introduce the idea behind variational inference. Recall that
525 any Bayesian inference problem deals with the joint distribution between observations \underline{x} and
526 unobserved latent variables $\underline{\theta}$. This joint distribution can be written as the product of a
527 distribution of the observations \underline{x} conditioned on the $\underline{\theta}$ and the marginal distribution of these
528 latent variables, i.e.,

$$\pi(\underline{x}, \underline{\theta}) = \pi(\underline{x} | \underline{\theta})\pi(\underline{\theta}). \quad (\text{S1})$$

529 A Bayesian inference pipeline's objective is to compute the latent variables' posterior probability
530 given a set of observations. This computation is equivalent to updating our prior beliefs
531 about the set of values that the latent variables take after taking in new data. We write this
532 as Bayes theorem

$$\pi(\underline{\theta} | \underline{x}) = \frac{\pi(\underline{x} | \underline{\theta})\pi(\underline{\theta})}{\pi(\underline{x})}. \quad (\text{S2})$$

533 The main technical challenge for working with Equation S2 comes from the computation
534 of the denominator, also known as the *evidence* or the *marginalized likelihood*. The reason
535 computing this term is challenging is because it involves a (potentially) high-dimensional
536 integral of the form

$$\pi(\underline{x}) = \int \cdots \int d^K \underline{\theta} \pi(\underline{x}, \underline{\theta}) = \int \cdots \int d^K \underline{\theta} \pi(\underline{x} | \underline{\theta})\pi(\underline{\theta}), \quad (\text{S3})$$

537 where K is the dimensionality of the $\underline{\theta}$ vector. Here, the integrals are taken over the support—
538 the set of values valid for the distribution—of $\pi(\underline{\theta})$. However, only a few selected distributions
539 have a closed analytical form; thus, in most cases Equation S3 must be solved numerically.

540 Integration in high-dimensional spaces can be computationally extremely challenging. For a
 541 naive numerical quadrature procedure, integrating over a grid of values for each dimension of
 542 θ comes with an exponential explosion of the number of required grid point evaluations, most
 543 of which do not contribute significantly to the integration. To gain visual intuition about
 544 this challenge, imagine integrating the function depicted in Figure S1. If the location of the
 545 high-density region (dark peak) is unknown, numerical quadrature requires many grid points
 546 to ensure we capture this peak. However, most of the numerical evaluations of the function on
 547 the grid points do not contribute significantly to the integral. Therefore, our computational
 548 resources are wasted on insignificant evaluations. This only gets worse as the number of
 549 dimensions increases since the number of grid point evaluation scales exponentially.

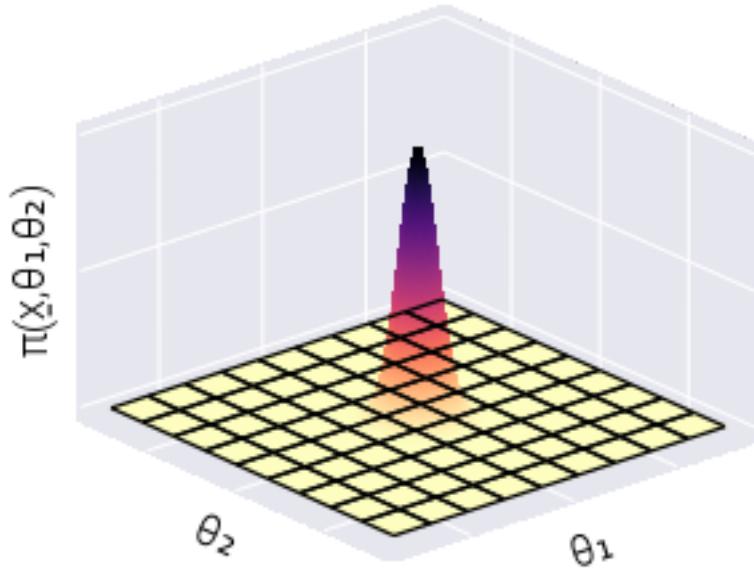


Figure S1. High-dimensional numerical quadrature does not scale with dimensionality.

Schematic depiction of the problem with naive numerical quadrature to integrate over an unknown density. While the density is concentrated on the dark peak, most of the evaluations over the $x_1 - x_2$ grid do not contribute to the value of the integral

550 Modern Markov Chain Monte Carlo algorithms, such as Hamiltonian Monte Carlo, can
 551 efficiently perform this high-dimensional integration by utilizing gradient information from
 552 the target density Betancourt [6]. Nevertheless, these sampling-based methods become
 553 prohibitively slow for the number of dimensions our present inference problem presents. Thus,
 554 there is a need to find scalable methods for the inference problem in Equation S2.

555 Variational inference circumvents these technical challenges by proposing an approximate

556 solution to the problem. Instead of working with the posterior distribution in its full glory
 557 $\pi(\underline{\theta} | \underline{x})$, let us propose an approximate posterior distribution q_ϕ that belongs to a distribution
 558 family fully parametrized by ϕ . For example, let us say that the distribution q_ϕ belongs
 559 to the family of multivariate Normal distributions such that $\phi = (\underline{\mu}, \underline{\Sigma})$, where $\underline{\mu}$ is the
 560 vector of means and $\underline{\Sigma}$ is the covariance matrix. If we replace π by q_ϕ , we want q_ϕ to
 561 resemble the original posterior as much as possible. Mathematically, this can be expressed as
 562 minimizing a “*distance metric*”—the Kullback-Leibler (KL) divergence, for example—between
 563 the distributions. Note that we use quotation marks because, formally, the KL divergence is
 564 not a distance metric since it is not symmetric. Nevertheless, the variational objective is set
 565 to find a distribution q_ϕ^* such that

$$q_\phi^*(\underline{\theta}) = \min_{\phi} D_{KL}(q_\phi(\underline{\theta}) || \pi(\underline{\theta} | \underline{x})), \quad (\text{S4})$$

566 where D_{KL} is the KL divergence. Furthermore, we highlight that the KL divergence is a
 567 strictly positive number, i.e.,

$$D_{KL}(q_\phi(\underline{\theta}) || \pi(\underline{\theta} | \underline{x})) \geq 0, \quad (\text{S5})$$

568 as this property will become important later on.

569 At first sight, Equation S4 does not improve the situation but only introduces further technical
 570 complications. After all, the definition of the KL divergence

$$D_{KL}(q_\phi(\underline{\theta}) || \pi(\underline{\theta} | \underline{x})) \equiv \int \cdots \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln \frac{q_\phi(\underline{\theta})}{\pi(\underline{\theta} | \underline{x})}, \quad (\text{S6})$$

571 includes the posterior distribution $\pi(\underline{\theta} | \underline{x})$ we are trying to get around. However, let us
 572 manipulate Equation S6 to beat it to a more reasonable form. First, we can use the properties
 573 of the logarithms to write

$$D_{KL}(q_\phi(\underline{\theta}) || \pi(\underline{\theta} | \underline{x})) = \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln q_\phi(\underline{\theta}) - \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln \pi(\underline{\theta} | \underline{x}), \quad (\text{S7})$$

574 where, for convenience, we write a single integration sign ($d^K \underline{\theta}$ still represents a multi-
 575 dimensional differential). For the second term in Equation S7, we can substitute the term
 576 inside the logarithm using Equation S2. This results in

$$\begin{aligned} D_{KL}(q_\phi(\underline{\theta}) || \pi(\underline{\theta} | \underline{x})) &= \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln q_\phi(\underline{\theta}) \\ &\quad - \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln \left(\frac{\pi(\underline{x} | \underline{\theta}) \pi(\underline{\theta})}{\pi(\underline{x})} \right). \end{aligned} \quad (\text{S8})$$

577 Again, using the properties of logarithms, we can split Equation S8, obtaining

$$\begin{aligned} D_{KL}(q_\phi(\underline{\theta}) || \pi(\underline{\theta} | \underline{x})) &= \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln q_\phi(\underline{\theta}) \\ &\quad - \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln \pi(\underline{x} | \underline{\theta}) \\ &\quad - \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln \pi(\underline{\theta}) \\ &\quad + \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln \pi(\underline{x}). \end{aligned} \quad (\text{S9})$$

578 It is convenient to write Equation S9 as

$$\begin{aligned} D_{KL}(q_\phi(\underline{\theta}) \parallel \pi(\underline{\theta} \mid \underline{x})) &= \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln \frac{q_\phi(\underline{\theta})}{\pi(\underline{\theta})} \\ &\quad - \int d^K \underline{\theta} q_\phi(\underline{\theta}) \ln \pi(\underline{x} \mid \underline{\theta}) \\ &\quad + \ln \pi(\underline{x}) \int d^K \underline{\theta} q_\phi(\underline{\theta}), \end{aligned} \quad (\text{S10})$$

579 where for the last term, we can take $\ln \pi(\underline{x})$ out of the integral since it does not depend on
580 $\underline{\theta}$. Lastly, we utilize two properties:

581 1. The proposed approximate distribution must be normalized, i.e.,

$$\int d^K \underline{\theta} q_\phi(\underline{\theta}) = 1. \quad (\text{S11})$$

582 2. The law of the unconscious statistician (LOTUS) establishes that for any probability
583 density function, it must be true that

$$\int d^K \underline{\theta} q_\phi(\underline{\theta}) f(\underline{\theta}) = \langle f(\underline{\theta}) \rangle_{q_\phi}, \quad (\text{S12})$$

584 where $\langle \cdot \rangle_{q_\phi}$ is the expected value over the q_ϕ distribution.

585 Using these two properties, the positivity constraint on the KL divergence in Equation S5,
586 and the definition of the KL divergence in Equation S6 we can rewrite Equation S10 as

$$D_{KL}(q_\phi(\underline{\theta}) \parallel \pi(\underline{\theta})) - \langle \ln \pi(\underline{x} \mid \underline{\theta}) \rangle_{q_\phi} \geq -\ln \pi(\underline{x}). \quad (\text{S13})$$

587 Multiplying by a minus one, we have the functional form of the so-called evidence lower
588 bound (ELBO) Kingma and Welling [9],

$$\underbrace{\ln \pi(\underline{x})}_{\text{log evidence}} \geq \underbrace{\langle \ln \pi(\underline{x} \mid \underline{\theta}) \rangle_{q_\phi} - D_{KL}(q_\phi(\underline{\theta}) \parallel \pi(\underline{\theta}))}_{\text{ELBO}}. \quad (\text{S14})$$

589 Let us recapitulate where we are. We started by presenting the challenge of working with
590 Bayes' theorem, as it requires a high-dimensional integral of the form in Equation S3. As an
591 alternative, variational inference posits to approximate the posterior distribution $\pi(\underline{\theta} \mid \underline{x})$ with
592 a parametric distribution $q_\phi(\underline{\theta})$. By minimizing the KL divergence between these distributions,
593 we arrive at the result in Equation S14, where the left-hand side—the log marginalized
594 likelihood or log evidence—we cannot compute for technical/computational reasons. However,
595 the right-hand side is composed of things we can easily evaluate. We can easily evaluate
596 the log-likelihood $\ln \pi(\underline{x} \mid \underline{\theta})$ and the KL divergence between our proposed approximate
597 distribution $q_\phi(\underline{\theta})$ and the prior distribution $\pi(\underline{\theta})$. Moreover, we can compute the gradients of
598 these functions with respect to the parameters of our proposed distribution. This last point
599 implies that we can change the parameters of the proposed distribution to maximize the

600 ELBO. And, although we cannot compute the left-hand side of Equation S14, we know that
 601 however large we make the ELBO, it will always be smaller than (or equal) the log-marginal
 602 likelihood. Therefore, the larger we can make the ELBO by modifying the parameters ϕ , the
 603 closer it gets to the log-marginal likelihood, and, as a consequence, the better our proposed
 604 distribution $q_\phi(\underline{\theta})$ gets to the true posterior distribution $\pi(\underline{\theta} | \underline{x})$.

605 In this sense, variational inference turns the intractable numerical integration problem to an
 606 optimization routine, for which there are several algorithms available.

607 **ADVI algorithm**

608 To maximize the right-hand side of Equation S14, the Automatic Differentiation Variational
 609 Inference (ADVI) algorithm developed in⁷ takes advantage of advances in probabilistic
 610 programming languages to generate a robust method to perform this optimization. Without
 611 going into the details of the algorithm implementation, for our purposes, it suffices to say
 612 that we define our joint distribution $\pi(\underline{\theta}, \underline{x})$ as the product defined in Equation S1. ADVI
 613 then proposes an approximate variational distribution q_ϕ that can either be a multivariate
 614 Normal distribution with a diagonal covariance matrix, i.e.,

$$\phi = (\underline{\mu}, \underline{\underline{D}}), \quad (\text{S15})$$

615 where $\underline{\underline{D}}$ is the identity matrix, with the diagonal elements given by the vector of variances
 616 $\underline{\sigma}^2$ for each variable or a full-rank multivariate Normal distribution

$$\phi = (\underline{\mu}, \underline{\underline{\Sigma}}). \quad (\text{S16})$$

617 Then, the parameters are initialized in some value ϕ_o . These parameters are iteratively
 618 updated by computing the gradient of the ELBO (right-hand side of Equation S14), hereafter
 619 defined as \mathcal{L} , with respect to the parameters,

$$\nabla_\phi \mathcal{L} = \nabla_{\underline{\mu}} \mathcal{L} + \nabla_{\underline{\sigma}} \mathcal{L}, \quad (\text{S17})$$

620 and then computing

$$\phi_{t+1} = \phi_t + \eta \nabla_\phi \mathcal{L},$$

621 where η defines the step size.

622 This short explanation behind the ADVI algorithm is intended only to gain intuition on
 623 how the optimal variational distribution q_ϕ be computed. There are many nuances in the
 624 implementation of the ADVI algorithm. We invite the reader to look at the original reference
 625 for further details.

626 **Defining the Bayesian model**

627 In the main text, we specify the inference problem we must solve as being of the form

$$\pi(\underline{s}^M, \bar{s}_T, \underline{F} | \underline{R}) \propto \pi(\underline{R} | \underline{s}^M, \bar{s}_T, \underline{F}) \pi(\underline{s}^M, \bar{s}_T, \underline{F}). \quad (\text{S18})$$

628 Here, we briefly define the missing nuisance parameters. Let

$$\bar{s}_T = (\bar{s}_1, \bar{s}_2, \dots, \bar{s}_{T-1})^\dagger, \quad (\text{S19})$$

629 be the vector containing the $T - 1$ population mean fitness we compute from the T time
630 points where measurements were taken. We have $T - 1$ since the value of any \bar{s}_t requires
631 cycle numbers t and $t + 1$. Furthermore, let the matrix \underline{F} be a $T \times B$ matrix containing all
632 frequency values. As with Equation 12 in the main text, we can split \underline{F} into two matrices of
633 the form

$$\underline{F} = \begin{bmatrix} \underline{F}^N & \underline{F}^M \end{bmatrix}, \quad (\text{S20})$$

634 to separate the corresponding neutral and non-neutral barcode frequencies.

635 Let us now define each of the terms in Equation 18 described in Section of the main text.
636 The following sections will specify the functional form each of these terms takes.

637 **Frequency uncertainty** $\pi(\underline{F} | \underline{R})$

638 We begin with the probability of the frequency values given the raw barcode reads. The first
639 assumption is that the inference of the frequency values for time t is independent of any other
640 time. Therefore, we can write the joint probability distribution as a product of independent
641 distributions of the form

$$\pi(\underline{F} | \underline{R}) = \prod_{t=1}^T \pi(\underline{f}_t | \underline{r}_t), \quad (\text{S21})$$

642 where \underline{f}_t and \underline{r}_t are the t -th row of the matrix containing all of the measurements for time t .
643 We imagine that when the barcode reads are obtained via sequencing, the quantified number
644 of reads is a Poisson sample from the “true” underlying number of barcodes within the pool.
645 This translates to assuming that the number of reads for each barcode at any time point $r_t^{(b)}$
646 is an independent Poisson random variable, i.e.,

$$r_t^{(b)} \sim \text{Poiss}(\lambda_t^{(b)}), \quad (\text{S22})$$

647 where the symbol “~” is read “distributed as.” Furthermore, for a Poisson distribution, we
648 have that

$$\lambda_t^{(b)} = \langle r_t^{(b)} \rangle = \left\langle \left(r_t^{(b)} - \langle r_t^{(b)} \rangle \right)^2 \right\rangle, \quad (\text{S23})$$

649 where $\langle \cdot \rangle$ is the expected value. In other words the Poisson parameter is equal to the mean
650 and variance of the distribution. The Poisson distribution has the convenient property that

651 for two Poisson distributed random variables $X \sim \text{Poiss}(\lambda_x)$ and $Y \sim \text{Poiss}(\lambda_y)$, we have
 652 that

$$Z \equiv X + Y \sim \text{Poiss}(\lambda_x + \lambda_y). \quad (\text{S24})$$

653 This additivity allows us to write the total number of reads at time t n_t also as a Poisson-
 654 distributed random variable of the form

$$n_t \sim \text{Poiss} \left(\sum_{b=1}^B \lambda_t^{(b)} \right), \quad (\text{S25})$$

655 where the sum is taken over all B barcodes.

656 If the total number of reads is given by Equation S25, the array with the number of reads for
 657 each barcode at time t , \underline{r}_t is then distributed as

$$\underline{r}_t \sim \text{Multinomial}(n_t, \underline{f}_t), \quad (\text{S26})$$

658 where each of the B entries of the frequency vector \underline{f}_t is a function of the $\underline{\lambda}_t$ vector, given
 659 by

$$f_t^{(b)} \equiv f_t^{(b)}(\underline{\lambda}_t) = \frac{\lambda_t^{(b)}}{\sum_{b'=1}^B \lambda_t^{(b')}}. \quad (\text{S27})$$

660 In other words, we can think of the B barcode counts as independent Poisson samples or as a
 661 single multinomial draw with a random number of total draws, n_t , and the frequency vector
 662 \underline{f}_t we are interested in. Notice that Equation S27 is a deterministic function that connects
 663 the Poisson parameters to the frequencies. Therefore, we have the equivalence that

$$\pi(\underline{f}_t | \underline{r}_t) = \pi(\underline{\lambda}_t | \underline{r}_t), \quad (\text{S28})$$

664 meaning that the uncertainty comes from the $\underline{\lambda}_t$ vector. By Bayes theorem, we therefore
 665 write

$$\pi(\underline{\lambda}_t | n_t, \underline{r}_t) \propto \pi(n_t, \underline{r}_t | \underline{\lambda}_t) \pi(\underline{\lambda}_t), \quad (\text{S29})$$

666 where we explicitly include the dependence on n_t . This does not affect the distribution or
 667 brings more uncertainty because \underline{r}_t already contains all the information to compute n_t since

$$n_t = \sum_{b=1}^B r_t^{(b)}. \quad (\text{S30})$$

668 But adding the variable allows us to factorize Equation S29 as

$$\pi(\underline{\lambda}_t | n_t, \underline{r}_t) \propto \pi(\underline{r}_t | n_t, \underline{\lambda}_t) \pi(n_t | \underline{\lambda}_t) \pi(\underline{\lambda}_t) \quad (\text{S31})$$

669 We then have

$$\underline{r}_t | n_t, \underline{\lambda}_t \sim \text{Multinomial}(n_t, \underline{f}_t(\underline{\lambda}_t)). \quad (\text{S32})$$

670 Furthermore, we have

$$n_t | \underline{\lambda}_t \sim \text{Poiss} \left(\sum_{b=1}^B \lambda_t^{(b)} \right).$$

671 {#eq=freq_n_bayes} Finally, for our prior $\pi(\underline{\lambda}_t)$, we first assume each parameter is indepen-
672 dent, i.e.,

$$\pi(\underline{\lambda}_t) = \prod_{b=1}^B \pi(\lambda_t^{(b)}).$$

673 A reasonable prior for each $\lambda_t^{(b)}$ representing the expected number of reads for barcode b
674 should span several orders of magnitude. Furthermore, we assume that no barcode in the
675 dataset ever goes extinct. Thus, no frequency can equal zero, facilitating the computation
676 of the log frequency ratios needed to infer the relative fitness. The log-normal distribution
677 satisfies these constraints; therefore, for the prior, we assume

$$\lambda_t^{(b)} \sim \log \mathcal{N}(\mu_{\lambda_t^{(b)}}, \sigma_{\lambda_t^{(b)}}), \quad (\text{S33})$$

678 with $\mu_{\lambda_t^{(b)}}, \sigma_{\lambda_t^{(b)}}$ as the user-defined parameters that characterize the prior distribution.

679 Summary

680 Putting all the pieces developed in this section together gives a term for our inference of the
681 form

$$\pi(\underline{F} | \underline{R}) \propto \prod_{t=1}^T \left\{ \pi(r_t | n_t, \underline{\lambda}_t) \pi(n_t | \underline{\lambda}_t) \left[\prod_{b=1}^B \pi(\lambda_t^{(b)}) \right] \right\} \quad (\text{S34})$$

682 where

$$r_t | n_t, \underline{\lambda}_t \sim \text{Multinomial}(n_t, f_t(\underline{\lambda}_t)), \quad (\text{S35})$$

$$n_t | \underline{\lambda}_t \sim \text{Poiss} \left(\sum_{b=1}^B \lambda_t^{(b)} \right). \quad (\text{S36})$$

684 and

$$\lambda_t^{(b)} \sim \log \mathcal{N}(\mu_{\lambda_t^{(b)}}, \sigma_{\lambda_t^{(b)}}), \quad (\text{S37})$$

685 Population mean fitness uncertainty $\pi(\bar{s}_T | \underline{F}, \underline{R})$

686 Next, we turn our attention to the problem of determining the population mean fitnesses \bar{s}_T .
687 First, we notice that our fitness model in Equation 3 does not include the value of the raw
688 reads. They enter the calculation indirectly through the inference of the frequency values we
689 developed in Section . This means that we can remove the conditioning of the value of \bar{s}_T
690 on the number of reads, obtaining a simpler probability function

$$\pi(\bar{s}_T | \underline{F}, \underline{R}) = \pi(\bar{s}_T | \underline{F}). \quad (\text{S38})$$

691 Moreover, our fitness model does not directly explain how the population mean fitness evolves
692 over time. In other words, our model cannot explicitly compute the population mean fitness
693 at time $t + 1$ from the information we have about time t . Given this model limitation, we are

694 led to assume that we must infer each \bar{s}_t independently. Expressing this for our inference
 695 results in

$$\pi(\underline{\bar{s}}_T \mid \underline{\underline{F}}) = \prod_{t=1}^{T-1} \pi(\bar{s}_t \mid \underline{f}_t, \underline{f}_{t+1}), \quad (\text{S39})$$

696 where we split our matrix $\underline{\underline{F}}$ for each time point and only kept the conditioning on the relevant
 697 frequencies needed to compute the mean fitness at time t .

698 Although our fitness model in Equation 3 also includes the relative fitness $s^{(m)}$, to infer the
 699 population mean fitness we only utilize data from the neutral lineages that, by definition,
 700 have a relative fitness $s^{(n)} = 0$. Therefore, the conditioning on Equation S39 can be further
 701 simplified by only keeping the frequencies of the neutral lineages, i.e.,

$$\pi(\bar{s}_t \mid \underline{f}_t, \underline{f}_{t+1}) = \pi(\bar{s}_t \mid \underline{f}_t^N, \underline{f}_{t+1}^N). \quad (\text{S40})$$

702 Recall that in Section we emphasized that the frequencies $f_t^{(n)}$ do not represent the true
 703 frequency of a particular lineage in the population but rather a “normalized number of
 704 cells.” Therefore, it is safe to assume each of the N neutral lineages’ frequencies is changing
 705 independently. The correlation of how increasing the frequency of one lineage will decrease
 706 the frequency of others is already captured in the model presented in Section . Thus, we
 707 write

$$\pi(\bar{s}_t \mid \underline{f}_t^N, \underline{f}_{t+1}^N) = \prod_{n=1}^N \pi(\bar{s}_t \mid f_t^{(n)}, f_{t+1}^{(n)}). \quad (\text{S41})$$

708 Now, we can focus on one of the terms on the right-hand side of Equation S41. Writing
 709 Bayes theorem results in

$$\pi(\bar{s}_t \mid f_t^{(n)}, f_{t+1}^{(n)}) \propto \pi(f_t^{(n)}, f_{t+1}^{(n)} \mid \bar{s}_t) \pi(\bar{s}_t). \quad (\text{S42})$$

710 Notice the likelihood defines the joint distribution of neutral barcode frequencies conditioned
 711 on the population mean fitness. However, rewriting our fitness model in Equation 3 for a
 712 neutral lineage to leave frequencies on one side and fitness on the other results in

$$\frac{f_{t+1}^{(n)}}{f_t^{(n)}} = e^{-\bar{s}_t \tau}. \quad (\text{S43})$$

713 Equation S43 implies that our fitness model only relates **the ratio** of frequencies and not the
 714 individual values. To get around this complication, we define

$$\gamma_t^{(b)} \equiv \frac{f_{t+1}^{(b)}}{f_t^{(b)}}, \quad (\text{S44})$$

715 as the ratio of frequencies between two adjacent time points for any barcode b . This allows
 716 us to rewrite the joint distribution $\pi(f_t^{(n)}, f_{t+1}^{(n)} \mid \bar{s}_t)$ as

$$\pi(f_t^{(n)}, f_{t+1}^{(n)} \mid \bar{s}_t) = \pi(f_t^{(n)}, \gamma_t^{(n)} \mid \bar{s}_t). \quad (\text{S45})$$

717 Let us rephrase this subtle but necessary change of variables since it is a key part of the
 718 inference problem: our series of independence assumptions lead us to Equation S42 that
 719 relates the value of the population mean fitness \bar{s}_t to the frequency of a neutral barcode at
 720 times t and $t + 1$. However, as shown in Equation S43, our model functionally relates the
 721 ratio of frequencies—that we defined as $\gamma_t^{(n)}$ —and not the independent frequencies to the
 722 mean fitness. Therefore, instead of writing for the likelihood the joint distribution of the
 723 frequency values at times t and $t + 1$ conditioned on the mean fitness, we write the joint
 724 distribution of the barcode frequency at time t and the ratio of the frequencies. These **must**
 725 **be** equivalent joint distributions since there is a one-to-one mapping between $\gamma_t^{(n)}$ and $f_{t+1}^{(n)}$
 726 for a given value of $f_t^{(n)}$. Another way to phrase this is to say that knowing the frequency at
 727 time t and at time $t + 1$ provides the same amount of information as knowing the frequency
 728 at time t and the ratio of the frequencies. This is because if we want to obtain $f_{t+1}^{(n)}$ given
 729 this information, we simply compute

$$f_{t+1}^{(n)} = \gamma_t^{(n)} f_t^{(n)}. \quad (\text{S46})$$

730 The real advantage of rewriting the joint distribution as in Equation S45 comes from splitting
 731 this joint distribution as a product of conditional distributions of the form

$$\pi(f_t^{(n)}, \gamma_t^{(n)} | \bar{s}_t) = \pi(f_t^{(n)} | \gamma_t^{(n)}, \bar{s}_t) \pi(\gamma_t^{(n)} | \bar{s}_t). \quad (\text{S47})$$

732 Written in this form, we can finally propose a probabilistic model for how the mean fitness
 733 relates to the frequency ratios we determine in our experiments. The second term on the
 734 right-hand side of Equation S47 relates how the determined frequency ratio $\gamma_t^{(b)}$ relates to
 735 the mean fitness \bar{s}_t . From Equation S43 and Equation S44, we can write

$$\ln \gamma_t^{(n)} = -\bar{s}_t + \varepsilon_t^{(n)}, \quad (\text{S48})$$

736 where, for simplicity, we set $\tau = 1$. Note that we added an extra term, $\varepsilon_t^{(n)}$, characterizing
 737 the deviations of the measurements from the theoretical model. We assume these errors are
 738 normally distributed with mean zero and some standard deviation σ_t , implying that

$$\ln \gamma_t^{(n)} | \bar{s}_t, \sigma_t \sim \mathcal{N}(-\bar{s}_t, \sigma_t), \quad (\text{S49})$$

739 where we include the nuisance parameter σ_t to be determined. If we assume the log frequency
 740 ratio is normally distributed, this implies the frequency ratio itself is distributed log-normal.
 741 This means that

$$\gamma_t^{(n)} | \bar{s}_t, \sigma_t \sim \log \mathcal{N}(-\bar{s}_t, \sigma_t). \quad (\text{S50})$$

742 Having added the nuisance parameter σ_t implies that we must update Equation S42 to

$$\pi(\bar{s}_t, \sigma_t | f_t^{(n)}, f_{t+1}^{(n)}) \propto \pi(f_t^{(n)}, \gamma_t^{(n)} | \bar{s}_t, \sigma_t) \pi(\bar{s}_t) \pi(\sigma_t), \quad (\text{S51})$$

743 where we assume the prior for each parameter is independent, i.e.,

$$\pi(\bar{s}_t, \sigma_t) = \pi(\bar{s}_t) \pi(\sigma_t). \quad (\text{S52})$$

744 For numerical stability, we will select weakly-informative priors for both of these parameters.
 745 In the case of the nuisance parameter σ_t , the prior must be restricted to positive values only,
 746 since standard deviations cannot be negative.

747 For the first term on the right-hand side of Equation S47, $\pi(f_t^{(n)} | \gamma_t^{(n)}, \bar{s}_t)$, we remove the
 748 conditioning on the population mean fitness since it does not add any information on top of
 749 what the frequency ratio $\gamma_t^{(n)}$ already gives. Therefore, we have

$$\pi(f_t^{(n)} | \gamma_t^{(n)}, \bar{s}_t) = \pi(f_t^{(n)} | \gamma_t^{(n)}). \quad (\text{S53})$$

750 The right-hand side of Equation S53 asks us to compute the probability of observing a
 751 frequency value $f_t^{(n)}$ given that we get to observe the ratio $\gamma_t^{(n)}$. If the ratio happened to be
 752 $\gamma_t^{(n)} = 2$, we could have $f_{t+1}^{(n)} = 1$ and $f_{t+1}^{(n)} = 0.5$, for example. Although, it would be equally
 753 likely that $f_{t+1}^{(n)} = 0.6$ and $f_{t+1}^{(n)} = 0.3$ or $f_{t+1}^{(n)} = 0.1$ and $f_{t+1}^{(n)} = 0.05$ for that matter. If we
 754 only get to observe the frequency ratio $\gamma_t^{(n)}$, we know that the numerator $f_{t+1}^{(n)}$ can only take
 755 values between zero and one, all of them being equally likely given only the information on
 756 the ratio. As a consequence, the value of the frequency in the denominator $f_t^{(n)}$ is restricted
 757 to fall in the range

$$f_t^{(n)} \in \left(0, \frac{1}{\gamma_t^{(n)}}\right]. \quad (\text{S54})$$

758 A priori, we do not have any reason to favor any value over any other, therefore it is natural
 759 to write

$$f_t^{(n)} | \gamma_t^{(n)} \sim \text{Uniform}\left(0, \frac{1}{\gamma_t^{(n)}}\right). \quad (\text{S55})$$

760 Summary

761 Putting all the pieces we have developed in this section together results in an inference for
 762 the population mean fitness values of the form

$$\pi(\bar{s}_T, \underline{\sigma}_T | \underline{F}) \propto \prod_{t=1}^{T-1} \left\{ \prod_{n=1}^N [\pi(f_t^{(n)} | \gamma_t^{(n)}) \pi(\gamma_t^{(n)} | \bar{s}_t, \sigma_t)] \pi(\bar{s}_t) \pi(\sigma_t) \right\}, \quad (\text{S56})$$

763 where we have

$$f_t^{(n)} | \gamma_t^{(n)} \sim \text{Uniform}\left(0, \frac{1}{\gamma_t^{(n)}}\right), \quad (\text{S57})$$

$$\gamma_t^{(n)} | \bar{s}_t, \sigma_t \sim \log \mathcal{N}(\bar{s}_t, \sigma_t), \quad (\text{S58})$$

$$\bar{s}_t \sim \mathcal{N}(0, \sigma_{\bar{s}_t}), \quad (\text{S59})$$

766 and

$$\sigma_t \sim \log \mathcal{N}(\mu_{\sigma_t}, \sigma_{\sigma_t}), \quad (\text{S60})$$

767 where $\sigma_{\bar{s}_t}$, μ_{σ_t} , and σ_{σ_t} are user-defined parameters.

768 **Mutant relative fitness uncertainty** $\pi(\underline{s}^M \mid \bar{s}_T, \underline{F}, \underline{R})$

769 The last piece of our inference is the piece that we care about the most: the probability
 770 distribution of all the mutants' relative fitness, given the inferred population mean fitness
 771 and the frequencies. First, we assume that all fitness values are independent of each other.
 772 This allows us to write

$$\pi(\underline{s}^M \mid \bar{s}_T, \underline{F}, \underline{R}) = \prod_{m=1}^M \pi(s^{(m)} \mid \bar{s}_T, \underline{F}, \underline{R}). \quad (\text{S61})$$

773 Furthermore, as was the case with the population mean fitness, our fitness model relates
 774 frequencies, not raw reads. Moreover, the fitness value of mutant m only depends on the
 775 frequencies of such mutant. Therefore, we can simplify the conditioning to

$$\pi(s^{(m)} \mid \bar{s}_T, \underline{F}, \underline{R}) = \pi(s^{(m)} \mid \bar{s}_T, \underline{f}^{(m)}), \quad (\text{S62})$$

776 where

$$\underline{f}^{(m)} = (f_0^{(m)}, f_1^{(m)}, \dots, f_T^{(m)})^\dagger, \quad (\text{S63})$$

777 is the vector containing the frequency time series for mutant m . Writing Bayes' theorem for
 778 the right-hand side of Equation S62 results in

$$\pi(s^{(m)} \mid \bar{s}_T, \underline{f}^{(m)}) \propto \pi(\underline{f}^{(m)} \mid \bar{s}_T, s^{(m)}) \pi(s^{(m)} \mid \bar{s}_T). \quad (\text{S64})$$

779 Notice the conditioning on the mean fitness values \bar{s}_T is not inverted since we already inferred
 780 these values.

781 Following the logic used in Section , let us define

$$\underline{\gamma}^{(m)} = (\gamma_0^{(m)}, \gamma_1^{(m)}, \dots, \gamma_{T-1}^{(m)})^\dagger, \quad (\text{S65})$$

782 where each entry $\gamma_t^{(m)}$ is defined by Equation S44. In the same way we rewrote the joint
 783 distribution between two adjacent time point frequencies to the joint distribution between
 784 one of the frequencies and the ratio of both frequencies in Equation S45, we can rewrite the
 785 joint distribution of the frequency time series for mutant m as

$$\pi(\underline{f}^{(m)} \mid \bar{s}_T, s^{(m)}) = \pi(f_0^{(m)}, \underline{\gamma}^{(m)} \mid \bar{s}_T, s^{(m)}). \quad (\text{S66})$$

786 One can think about Equation S66 as saying that knowing the individual frequencies at each
 787 time point contain equivalent information as knowing the initial frequency and the subsequent
 788 ratios of frequencies. This is because if we want to know the value of $f_1^{(m)}$ given the ratios,
 789 we only need to compute

$$f_1^{(m)} = \gamma_0^{(m)} f_0^{(m)}. \quad (\text{S67})$$

790 Moreover, if we want to know $f_2^{(m)}$, we have

$$f_2^{(m)} = \gamma_1^{(m)} f_1^{(m)} = \gamma_1^{(m)} (\gamma_0^{(m)} f_0^{(m)}), \quad (\text{S68})$$

791 and so on. We can then write the joint distribution on the right-hand side of Equation S66
 792 as a product of conditional distributions of the form

$$\begin{aligned} \pi(f_0^{(m)}, \underline{\gamma}^{(m)} | \bar{s}_T, s^{(m)}) &= \pi(f_0^{(m)} | \gamma_0^{(m)}, \dots, \gamma_{T-1}^{(m)}, \bar{s}_T, s^{(m)}) \times \\ &\quad \pi(\gamma_0^{(m)} | \gamma_1^{(m)}, \dots, \gamma_{T-1}^{(m)}, \bar{s}_T, s^{(m)}) \times \\ &\quad \pi(\gamma_1^{(m)} | \gamma_2^{(m)}, \dots, \gamma_{T-1}^{(m)}, \bar{s}_T, s^{(m)}) \times \\ &\quad \vdots \\ &\quad \pi(\gamma_{T-2}^{(m)} | \gamma_{T-1}^{(m)}, \bar{s}_T, s^{(m)}) \times \\ &\quad \pi(\gamma_{T-1}^{(m)} | \bar{s}_T, s^{(m)}). \end{aligned} \tag{S69}$$

793 Writing the fitness model in Equation 3 as

$$\gamma_t^{(m)} = \frac{f_{t+1}^{(m)}}{f_t^{(m)}} = e^{(s^{(m)} - s_t)\tau},$$

794 reveals that the value of each of the ratios $\gamma_t^{(m)}$ only depends on the corresponding fitness
 795 value \bar{s}_t and the relative fitness $s^{(m)}$. Therefore, we can remove most of the conditioning
 796 on the right-hand side of Equation S69, resulting in a much simpler joint distribution of the
 797 form

$$\begin{aligned} \pi(f_0^{(m)}, \underline{\gamma}^{(m)} | \bar{s}_T, s^{(m)}) &= \pi(f_0^{(m)} | \gamma_0^{(m)}) \times \\ &\quad \pi(\gamma_0^{(m)} | \bar{s}_0, s^{(m)}) \times \\ &\quad \pi(\gamma_1^{(m)} | \bar{s}_1, s^{(m)}) \times \\ &\quad \vdots \\ &\quad \pi(\gamma_{T-2}^{(m)} | \bar{s}_{T-2}, s^{(m)}) \times \\ &\quad \pi(\gamma_{T-1}^{(m)} | \bar{s}_{T-1}, s^{(m)}), \end{aligned} \tag{S70}$$

798 where for the first term on the right-hand side of Equation S70 we apply the same logic as in
 799 Equation S53 to remove all other dependencies. We emphasize that although Equation S70
 800 looks like a series of independent inferences, the value of the relative fitness $s^{(m)}$ is shared
 801 among all of them. This means that the parameter is not inferred individually for each time
 802 point, resulting in different estimates of the parameter, but each time point contributes
 803 independently to the inference of a single estimate of $s^{(m)}$.

804 Using equivalent arguments to those in Section , we assume

$$f_0^{(m)} | \gamma_0^{(m)} \sim \text{Uniform}\left(0, \frac{1}{\gamma_0^{(m)}}\right),$$

805 and

$$\gamma_t^{(m)} | \bar{s}_t, s^{(m)}, \sigma^{(m)} \sim \log \mathcal{N}(s^{(m)} - \bar{s}_t, \sigma^{(m)}), \tag{S71}$$

806 where we add the nuisance parameter $\sigma^{(m)}$ to the inference. Notice that this parameter is not
 807 indexed by time. This means that we assume the deviations from the theoretical prediction
 808 do not depend on time, but only on the mutant. Adding the nuisance parameter demands us
 809 to update Equation S64 to

$$\pi(s^{(m)}, \sigma^{(m)} | \bar{s}_T, \underline{f}^{(m)}) \propto \pi(\underline{f}^{(m)} | \bar{s}_T, s^{(m)}, \sigma^{(m)}) \pi(s^{(m)}) \pi(\sigma^{(m)}), \quad (\text{S72})$$

810 where we assume independent priors for both parameters. We also removed the conditioning
 811 on the values of the mean fitness as knowing such values does not change our prior information
 812 about the possible range of values that the parameters can take. As with the priors on
 813 Section , we will assign weakly-informative priors to these parameters.

814 **Summary**

815 With all pieces in place, we write the full inference of the relative fitness values as

$$\pi(\underline{s}^M, \underline{\sigma}^M | \bar{s}_T, \underline{F}) \propto \prod_{m=1}^M \left\{ \pi(f_0^{(m)} | \gamma_0^{(m)}) \prod_{t=0}^{T-1} [\pi(\gamma_t^{(m)} | \bar{s}_t, s^{(m)}, \sigma^{(m)})] \pi(s^{(m)}) \pi(\sigma^{(m)}) \right\}, \quad (\text{S73})$$

816 where

$$f_0^{(m)} | \gamma_0^{(m)} \sim \text{Uniform}\left(0, \frac{1}{\gamma_0^{(m)}}\right), \quad (\text{S74})$$

$$\gamma_t^{(m)} | \bar{s}_t, s^{(m)}, \sigma^{(m)} \sim \log \mathcal{N}(s^{(m)} - \bar{s}_t, \sigma^{(m)}), \quad (\text{S75})$$

$$s^{(m)} \sim \mathcal{N}(0, \sigma_{s^{(m)}}), \quad (\text{S76})$$

819 and

$$\sigma^{(m)} \sim \log \mathcal{N}(\mu_{\sigma^{(m)}}, \sigma_{\sigma^{(m)}}), \quad (\text{S77})$$

820 where $\sigma_{s^{(m)}}, \mu_{\sigma^{(m)}},$ and $\sigma_{\sigma^{(m)}}$ are user-defined parameters.

821 **Hierarchical models for multiple experimental replicates**

822 As detailed in Section of the main text, we define a Bayesian hierarchical model to analyze data
 823 from multiple experimental replicates. The implementation requires only slightly modifying
 824 the base model detailed in the previous sections. The hierarchical model defines a hyper-
 825 fitness parameter $\theta^{(m)}$ for every non-neutral barcode. We can thus collect all of the M
 826 hyperparameters in an array of the form

$$\underline{\theta}^M = (\theta^{(1)}, \dots, \theta^{(M)})^\dagger. \quad (\text{S78})$$

827 Our data now consists of a series of matrices $\underline{\underline{R}}_{[j]}$, where the subindex $[j]$ refers to the j -th
 828 experimental replicate. These matrices need not have the same number of rows, as the time

829 points measured for each replicate can vary. The statistical model we must define is then of
 830 the form

$$\begin{aligned}\pi(\underline{\theta}^M, \{\underline{s}_{[j]}^M\}, \{\bar{s}_{T[j]}\}, \{\underline{F}_{[j]}\} \mid \{\underline{R}_{[j]}\}) &\propto \pi(\{\underline{R}_{[j]}\} \mid \underline{\theta}^M, \{\underline{s}_{[j]}^M\}, \{\bar{s}_{T[j]}\}, \{\underline{F}_{[j]}\}) \times \\ &\quad \pi(\underline{\theta}^M, \{\underline{s}_{[j]}^M\}, \{\bar{s}_{T[j]}\}, \{\underline{F}_{[j]}\})\end{aligned}\quad (\text{S79})$$

831 where the parameters within curly braces with subindex $[j]$ indicate one set of parameters per
 832 experimental replicate. For example,

$$\{\underline{s}_{[j]}^M\} = \{\underline{s}_{[1]}^M, \underline{s}_{[2]}^M, \dots, \underline{s}_{[E]}^M\}, \quad (\text{S80})$$

833 where E is the number of experimental replicates.

834 Given the dependencies between the variables, we can factorize Equation S79 to be of the
 835 form

$$\begin{aligned}\pi(\underline{\theta}^M, \{\underline{s}_{[j]}^M\}, \{\bar{s}_{T[j]}\}, \{\underline{F}_{[j]}\} \mid \{\underline{R}_{[j]}\}) &= \pi(\underline{\theta}^M, \{\underline{s}_{[j]}^M\} \mid \{\bar{s}_{T[j]}\}, \{\underline{F}_{[j]}\}) \times \\ &\quad \pi(\{\bar{s}_{T[j]}\} \mid \{\underline{F}_{[j]}\}) \times \\ &\quad \pi(\{\underline{F}_{[j]}\} \mid \{\underline{R}_{[j]}\})\end{aligned}\quad (\text{S81})$$

836 Furthermore, the hierarchical structure only connects the replicates via the relative fitness
 837 parameters. This means that the population mean fitness values and the frequencies can
 838 be independently inferred for each dataset. This allows us to rewrite the right-hand side of
 839 Equation S81 as

$$\begin{aligned}\pi(\underline{\theta}^M, \{\underline{s}_{[j]}^M\}, \{\bar{s}_{T[j]}\}, \{\underline{F}_{[j]}\} \mid \{\underline{R}_{[j]}\}) &= \pi(\underline{\theta}^M, \{\underline{s}_{[j]}^M\} \mid \{\bar{s}_{T[j]}\}, \{\underline{F}_{[j]}\}) \times \\ &\quad \prod_{j=1}^E \left[\pi(\bar{s}_{T[j]} \mid \underline{F}_{[j]}) \pi(\underline{F}_{[j]} \mid \underline{R}_{[j]}) \right].\end{aligned}\quad (\text{S82})$$

840 The terms inside the square brackets in Equation S82 take the same functional form as those
 841 derived in Section and Section . Therefore, to implement the desired hierarchical model, we
 842 only need to focus on the first term on the right-hand side of Equation S82. A way to think
 843 about the structure of the hierarchical model is as follows: imagine each genotype as a “*true*”
 844 relative fitness value. However, every time we perform an experiment, small variations in the
 845 biotic and abiotic conditions—also known as batch effects—might result in small deviations
 846 from this value. We model this by defining a distribution for the hyper-fitness parameter—the
 847 ground truth we are interested in—and having each experimental replicate sample from
 848 this hyper-parameter distribution to determine the “*local*” fitness value. The wider the
 849 hyper-parameter distribution is the more variability between experimental replicates.

850 Writing Bayes’ theorem for the first term in Equation S82 results in

$$\pi(\underline{\theta}^M, \{\underline{s}_{[j]}^M\} \mid \{\underline{F}_{[j]}\}, \{\bar{s}_{T[j]}\}) \propto \pi(\{\underline{F}_{[j]}\} \mid \underline{\theta}^M, \{\underline{s}_{[j]}^M\}, \{\bar{s}_{T[j]}\}) \pi(\underline{\theta}^M, \{\underline{s}_{[j]}^M\} \mid \{\bar{s}_{T[j]}\}), \quad (\text{S83})$$

851 where we leave the conditioning on the population mean fitness as we did in Section . This
 852 expression can be simplified in two ways. First, the frequency values for each experimental
 853 replicate depend directly on the local fitness values and the corresponding population mean
 854 fitness, as the relationship between experimental replicates only occurs through the relative
 855 fitness values. Therefore, we can write

$$\pi(\underline{\theta}^M, \{s_{[j]}^M\} \mid \{\underline{F}_{[j]}\}, \{\bar{s}_{T[j]}\}) \propto \prod_{j=1}^E \left[\pi(\underline{F}_{[j]} \mid s_{[j]}^M, \bar{s}_{T[j]}) \right] \pi(\underline{\theta}^M, \{s_{[j]}^M\} \mid \{\bar{s}_{T[j]}\}). \quad (\text{S84})$$

856 Second, the relationship between the hyper-fitness and the local fitness values allows us to
 857 write their joint distribution as a conditional distribution where local fitness values depend on
 858 the global hyper-fitness value, obtaining

$$\pi(\underline{\theta}^M, \{s_{[j]}^M\} \mid \{\underline{F}_{[j]}\}, \{\bar{s}_{T[j]}\}) \propto \prod_{j=1}^E \left[\pi(F_{[j]} \mid s_{[j]}^M, \bar{s}_{T[j]}) \pi(s_{[j]}^M \mid \underline{\theta}^M) \right] \pi(\underline{\theta}^M). \quad (\text{S85})$$

859 Notice we removed the conditioning on the population mean fitness as our prior expectations
 860 of what the global hyper-fitness or local fitness value might be do not depend on these
 861 nuisance parameters.

862 The first term on the right-hand side of Equation S85 takes the same functional form as the
 863 one derived in Section . Therefore, all we are left with is to determine the functional forms
 864 for the hyper-prior $\pi(\underline{\theta}^M)$, and the conditional probability $\pi(s_{[j]}^M \mid \underline{\theta}^M)$. In analogy to the
 865 assumptions used for the fitness values in Section , we define the value of each hyper-fitness
 866 as independent. This means that we have

$$\pi(\underline{\theta}^M) = \prod_{m=1}^M \pi(\theta^{(m)}). \quad (\text{S86})$$

867 Furthermore, we assume this prior is of the form

$$\theta^{(m)} \sim \mathcal{N}(\mu_{\theta^{(m)}}, \sigma_{\theta^{(m)}}), \quad (\text{S87})$$

868 where $\mu_{\theta^{(m)}}$ and $\sigma_{\theta^{(m)}}$ are user-defined parameters encoding the prior expectations on the
 869 fitness values.

870 For the conditional distribution $\pi(s_{[j]}^M \mid \underline{\theta}^M)$, we use the so-called non-centered parametrization
 871 that avoids some of the intrinsic degeneracies associated with hierarchical models¹³. We
 872 invite the reader to check [this excellent blog](#) explaining the difficulties of working with
 873 hierarchical models. This non-centered parameterization implies that we introduce two
 874 nuisance parameters such that the local fitness $s_{[j]}^{(m)}$ is computed as

$$s_{[j]}^{(m)} = \theta^{(m)} + (\tau_{[j]}^{(m)} \times \xi_{[j]}^{(m)}), \quad (\text{S88})$$

875 where $\theta^{(m)}$ is the corresponding genotype hyper-fitness value, $\xi_{[j]}^{(m)}$ is a standard normal
 876 random variable, i.e.,

$$\xi_{[j]}^{(m)} \sim \mathcal{N}(0, 1), \quad (\text{S89})$$

877 that allows deviations from the hyper-fitness value to be either positive or negative, and $\tau_{[j]}^{(m)}$
878 is a strictly positive random variable that characterizes the deviation of the local fitness value
879 from the global hyper-fitness. We assume

$$\tau_{[j]}^{(m)} \sim \log \mathcal{N}(\mu_{\tau_{[j]}^{(m)}}, \sigma_{\tau_{[j]}^{(m)}}) \quad (\text{S90})$$

880 where $\mu_{\tau_{[j]}^{(m)}}$ and $\sigma_{\tau_{[j]}^{(m)}}$ are user-defined parameters capturing the expected magnitude of the
881 batch effects.

882 Defining prior probabilities

883 One aspect commonly associated—in both positive and negative ways—to Bayesian analysis
884 is the definition of prior probabilities. On the one hand, the naive textbook version of Bayesian
885 analysis defines the prior as encoding the information we have about the inference in question
886 before acquiring any data. This is the “ideal” use of priors that, whenever possible, should
887 be implemented. On the other hand, for most practitioners of Bayesian statistics in the age
888 of big data, the definition of prior becomes a tool to ensure the convergence of sampling
889 algorithms such as MCMC²⁵. However, for our particular problem, although we deal with
890 large amounts of data (inferences can be made for $> 10K$ barcodes over multiple time points,
891 resulting in $> 100K$ parameters), each barcode has very little data, as they are measured
892 only once per time point over < 10 growth-dilution cycles. Furthermore, it is incredibly
893 challenging to understand the noise sources related to culturing conditions, DNA extraction,
894 library preparation, etc., and encode them into reasonable prior distributions.

895 Empirically, our approach for this work defined the priors based solely on the neutral lineage
896 data, as they represent the only repeated measurements of a single genotype in our experi-
897 mental design. We acknowledge that defining the priors after observing the data might be
898 considered an incoherent inference. However, as expressed by Gelman et al. [25]

899 Incoherence is an unavoidable aspect of much real-world data analysis; and,
900 indeed, one might argue that as scientists we learn the most from the anomalies
901 and reassessments associated with episodes of incoherence.

902 With this in mind, we leave it to the reader to judge the selection of priors. Furthermore,
903 the software package associated with this work, BarBay.jl, is written so that users can
904 experiment with different prior selection criteria that fit their needs. We strongly advocate
905 that statistics should not be done in a black-box fit-all tool mindset but rather as a formal
906 way to encode the assumptions behind the analysis, subject to constructive criticism. With
907 this philosophical baggage behind us, let us now focus on how the priors used for this work
908 were selected.

909 **Naive neutral lineage-based priors**

910 For the base model presented in this work, the user-defined prior parameters include the
 911 following:

- 912 ■ Prior on population mean fitness (one per pair of adjacent time points)

$$\bar{s}_t \sim \mathcal{N}(\mu_{\bar{s}_t}, \sigma_{\bar{s}_t}). \quad (\text{S91})$$

- 913 ■ Prior on standard deviation associated with neutral lineages likelihood function (one
 914 per pair of adjacent time points)

$$\sigma_t \sim \log \mathcal{N}(\mu_{\sigma_t}, \sigma_{\sigma_t}). \quad (\text{S92})$$

- 915 ■ Prior on relative fitness (one per non-neutral barcode)

$$s^{(m)} \sim \mathcal{N}(\mu_{s^{(m)}}, \sigma_{s^{(m)}}). \quad (\text{S93})$$

- 916 ■ Prior on standard deviation associated with non-neutral lineages likelihood function
 917 (one per non-neutral barcode)

$$\sigma^{(m)} \sim \log \mathcal{N}(\mu_{\sigma^{(m)}}, \sigma_{\sigma^{(m)}}) \quad (\text{S94})$$

918 The BarBay.jl package includes a function `naive_prior` within the `stats` module. This
 919 function utilizes the data from the neutral lineages to determine some of the prior parameters
 920 to facilitate the inference algorithm's numerical convergence. In particular, it defines the
 921 population mean fitness parameter $\mu_{\bar{s}_t}$ as

$$\mu_{\bar{s}_t} = \frac{1}{N} \sum_{n=1}^N -\ln \left(\frac{r_{t+1}^{(n)}}{r_t^{(n)}} \right), \quad (\text{S95})$$

922 where N is the number of neutral lineages and $r_t^{(n)}$ is the number of neutral lineages. In
 923 other words, it defines the mean of the prior distribution as the mean of what one naively
 924 would compute from the neutral lineages, discarding cases where the ratio diverges because
 925 the denominator $r_t^{(n)} = 0$. For the variance parameter, we chose a value $\sigma_{\bar{s}_t} = 0.05$.

926 Furthermore, the `naive_prior` function defines the mean of the variance parameter as the
 927 standard deviation of the log frequency ratios for the neutral lineages, i.e.,

$$\mu_{\sigma_t} = \sqrt{\text{Var} \left(\frac{r_{t+1}^{(n)}}{r_t^{(n)}} \right)}, \quad (\text{S96})$$

928 where `Var` is the sample variance. This same value was utilized for the mean of the
 929 non-neutral barcode variance $\mu_{\sigma^{(m)}}$. While we assign the corresponding variances to be
 930 $\sigma_{\sigma_t} = \sigma_{\sigma^{(m)}} = 1$.

931 **Posterior predictive checks**

932 Throughout the main text, we allude to the concept of posterior predictive checks as a formal
 933 way to assess the accuracy of our inference pipeline. Here, we explain the mechanics behind
 934 the computation of these credible regions, given the output of the inference.

935 Bayesian models encode what is known as a *generative model*. This statement means that in
 936 our definition of the likelihood function and the prior distribution, we, as modelers, propose a
 937 mathematical function that captures all relevant relationships between unobserved (latent)
 938 variables. Therefore, when these latent variables are input into the mathematical model,
 939 this function *generates* data that should be, in principle, indistinguishable from the real
 940 observations if the model is a good account of the underlying processes involved in the
 941 phenomena of interest. This generative model implies that once we run the inference process
 942 and update our posterior beliefs about the state of the latent variables, we can input back the
 943 inferred values to our model and generate synthetic data. Furthermore, we can repeat this
 944 process multiple times to compute the range where we expect to observe our data conditioned
 945 on the accuracy of the model.

946 For our specific scenario, recall that our objective is to infer the relative fitness of a non-neutral
 947 lineage $s^{(m)}$ along with nuisance parameters related to the population mean fitness at each
 948 point, \bar{s}_t , and the barcode frequency time series $\underline{f}^{(m)}$. All these variables are related through
 949 our fitness model (see Section in the main text)

$$f_{t+1}^{(m)} = f_t^{(m)} e^{(s^{(m)} - s_t)\tau}. \quad (\text{S97})$$

950 As we saw, it is convenient to rewrite Equation S97 as

$$\frac{1}{\tau} \ln \frac{f_{t+1}^{(m)}}{f_t^{(m)}} = (s^{(m)} - s_t). \quad (\text{S98})$$

951 Written in this way, we separate the quantities we can compute from the experimental
 952 observations—the left-hand side of Equation S98 can be computed from the barcode reads—
 953 from the latent variables.

954 Although we perform the joint inference over all barcodes in the present work, let us focus on
 955 the inference task for a single barcode as if it were computed independently. For a non-neutral
 956 barcode, our task consists of computing the posterior probability

$$\pi(\theta | \underline{r}^{(m)}) = \pi(s^{(m)}, \sigma^{(m)}, \underline{s}_t, \underline{f}^{(m)} | \underline{r}^{(m)}), \quad (\text{S99})$$

957 where θ represents all parameters to be inferred and $\underline{r}^{(m)}$ is the vector with the barcode raw
 958 counts time series. The list of parameters are

- 959 ■ $s^{(m)}$: The barcode's relative fitness.
- 960 ■ $\sigma^{(m)}$: A nuisance parameter used in the likelihood to generate the data. This captures
 the expected deviation from Equation S98

- 962 ▪ \underline{s}_t : The vector with all population mean fitness for each pair of adjacent time points.
 963 ▪ $\underline{f}^{(m)}$: The vector with the barcode frequency time series.

964 Furthermore, let us define a naive estimate of the barcode frequency at time t as

$$\hat{f}_t^{(m)} = \frac{r_t^{(m)}}{\sum_{b=1}^B r_t^{(b)}}. \quad (\text{S100})$$

965 We can compute this quantity from the data by normalizing the raw barcode counts by
 966 the sum of all barcode counts. Furthermore, we can compute a naive estimate of the log
 967 frequency ratio from the raw barcode counts as

$$\ln \hat{\gamma}_t^{(m)} = \ln \frac{\hat{f}_{t+1}^{(m)}}{\hat{f}_t^{(m)}} \quad (\text{S101})$$

968 In our generative model, we assumed

$$\ln \gamma_t^{(m)} | \theta \sim \mathcal{N}(s^{(m)} - s_t, \sigma^{(m)}). \quad (\text{S102})$$

969 This implies that once we determine the posterior distribution of our parameters, we can
 970 generate synthetic values of $\ln \gamma_t^{(m)}$ that we can then compare with the values obtained from
 971 applying Equation S101 and Equation S101 to the raw data.

972 In practice, to compute the posterior predictive checks, we generate multiple samples from
 973 the posterior distribution $\pi(\theta | \underline{r}^{(m)})$

$$\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_N). \quad (\text{S103})$$

974 With these samples in hand, the BarBay.jl package includes the function `logfreq_ratio_bc_ppc`
 975 for non-neutral barcodes that uses this set of posterior parameter samples to generate
 976 samples from the distribution defined in Equation S102. For a large-enough number of
 977 samples, we can then compute the desired percentiles—5, 68, and 95 percentiles in all figures
 978 in the main text—that are equivalent to the corresponding credible regions. In other words,
 979 the range of values of $\ln \gamma_t^{(m)}$ generated by this bootstrap process can be used to compute
 980 the region where we expect to find our raw estimates $\ln \hat{\gamma}_t^{(m)}$ with the desired probability.
 981 The package BarBay.jl includes an equivalent function, `logfreq_ratio_popmean_ppc`,
 982 for neutral lineages.

983 Logistic growth simulation

984 In this section, we explain the simulations used to assess the validity of our inference pipeline.
 985 Let us begin by assuming that, since the strains are grown for two full days in the experiment,
 986 having left behind the exponential phase for almost an entire day, a simple exponential growth
 987 of the form

$$\frac{dn_i}{dt} = \lambda_i n_i, \quad (\text{S104})$$

988 where n_i is the number of cells of strain i , and λ_i is the corresponding growth rate is not
 989 enough. Instead, we will assume that the cells follow the logistic growth equation of the
 990 form

$$\frac{dn_i}{dt} = \lambda_i n_i \left(1 - \frac{\sum_{j=1}^N n_j}{\kappa}\right), \quad (\text{S105})$$

991 where κ is the carrying capacity, and N is the total number of strains in the culture.

992 The inference method is based on the model that assumes that the time passed between
 993 dilutions $\tau \approx 8$ generations, the change in frequency for a mutant barcode can be approximated
 994 from cycle t to the next cycle $t + 1$ as

$$f_{t+1}^{(m)} = f_t^{(m)} e^{(s^{(m)} - \bar{s}_t)\tau}, \quad (\text{S106})$$

995 where $s^{(m)}$ is the relative fitness for strain i compared to the ancestral strain and \bar{s}_t is
 996 the mean fitness of the population at cycle t . To test this assumption, we implemented
 997 a numerical experiment following the logistic growth model described in Equation S105.
 998 Figure S2 shows an example of the deterministic trajectories for 50 labeled neutral lineages
 999 and 1000 lineages of interest. The upper red curve that dominates the culture represents the
 1000 unlabeled ancestral strain included in the experimental design described in Section .

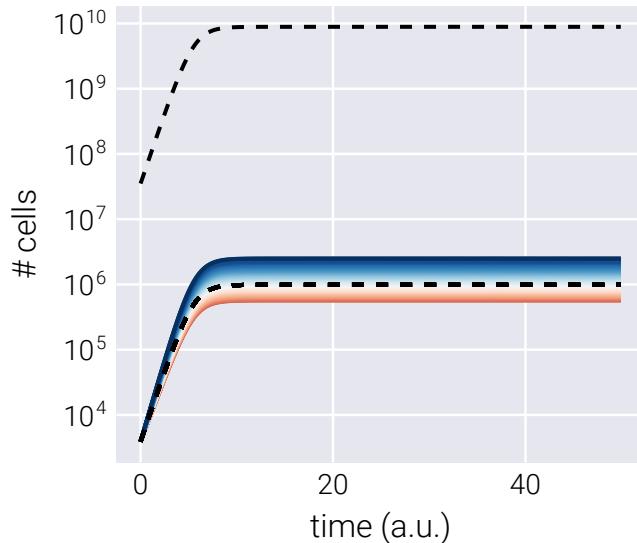


Figure S2. Logistic growth simulation over single growth cycle. The dashed line represents the neutral lineages, with the upper curve being the unlabeled neutral strain. Color curves represent the genotypes of interest colored by growth rate relative to the neutral lineage.

1001 To simulate multiple growth-dilution cycles, we take the population composition at the final
 1002 time point and use it to initialize a new logistic growth simulation. Figure S3 shows the
 1003 resulting number of cells at the last time point of a cycle over multiple growth-dilution cycles

1004 for the genotypes in @Figure S2. We can see that the adaptive lineages (blue curves) increase
 1005 in abundance, while detrimental lineages (red curves) decrease.

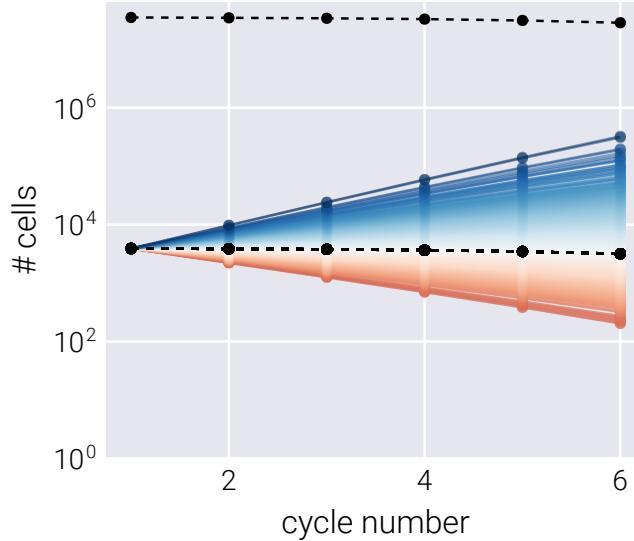


Figure S3. Growth-dilution cycles for logistic growth simulation. Each point represents the final number of cells after a growth cycle for each lineage. Colors are the same as in Figure S2.

1006 In Section , we derive the functional form to infer the relative fitness of each lineage as

$$\frac{1}{\tau} \ln \frac{f_{t+1}^{(b)}}{f_t^{(b)}} = (s^{(b)} - \bar{s}_t). \quad (\text{S107})$$

1007 Figure S4 shows the corresponding log frequency ratio curves for the logistic growth simulation.
 1008 The displacement of these curves with respect to the neutral lineages determines the ground
 1009 truth relative fitness value for these simulations.

1010 To simulate the experimental noise, we add two types of noise:

- 1011 1. Poisson noise between dilutions. For this, we take the final point of the logistic growth
- 1012 simulation and sample a random Poisson number based on this last point to set the
- 1013 initial condition for the next cycle.
- 1014 2. Gaussian noise when performing the measurements. When translating the underlying
- 1015 population composition to the number of reads, we can add a custom amount of
- 1016 Gaussian noise.

1017 Figure S5 shows the frequency trajectories (left panels) and log frequency ratios (right panels)
 1018 for a noiseless simulation (upper panels) and a simulation with added noise (lower panels).
 1019 The noiseless simulation is used to determine the relative fitness for each of the lineages,
 1020 which serves as the ground truth to be compared with the resulting inference.

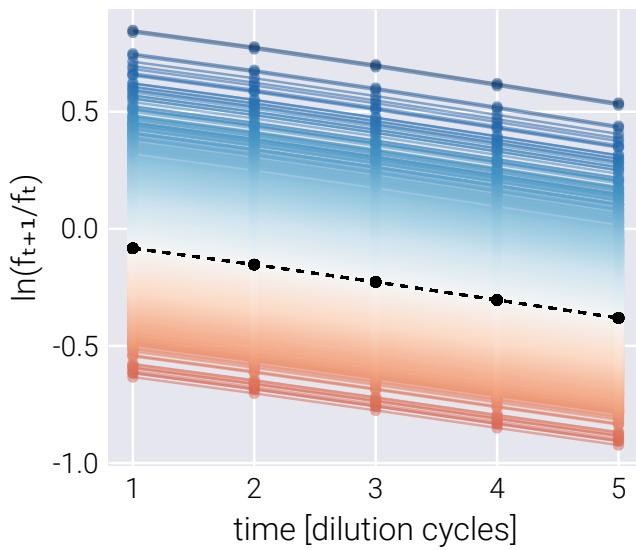


Figure S4. Log frequency ratio for logistic growth simulations. The relative distance of the color curves from the black dashed line determines the relative fitness of each lineage.

1021 Supplemental References

1022 ²⁵A. Gelman, D. Simpson, and M. Betancourt, “The Prior Can Often Only Be Understood in
 1023 the Context of the Likelihood”, [Entropy 19, 555 \(2017\)](#).

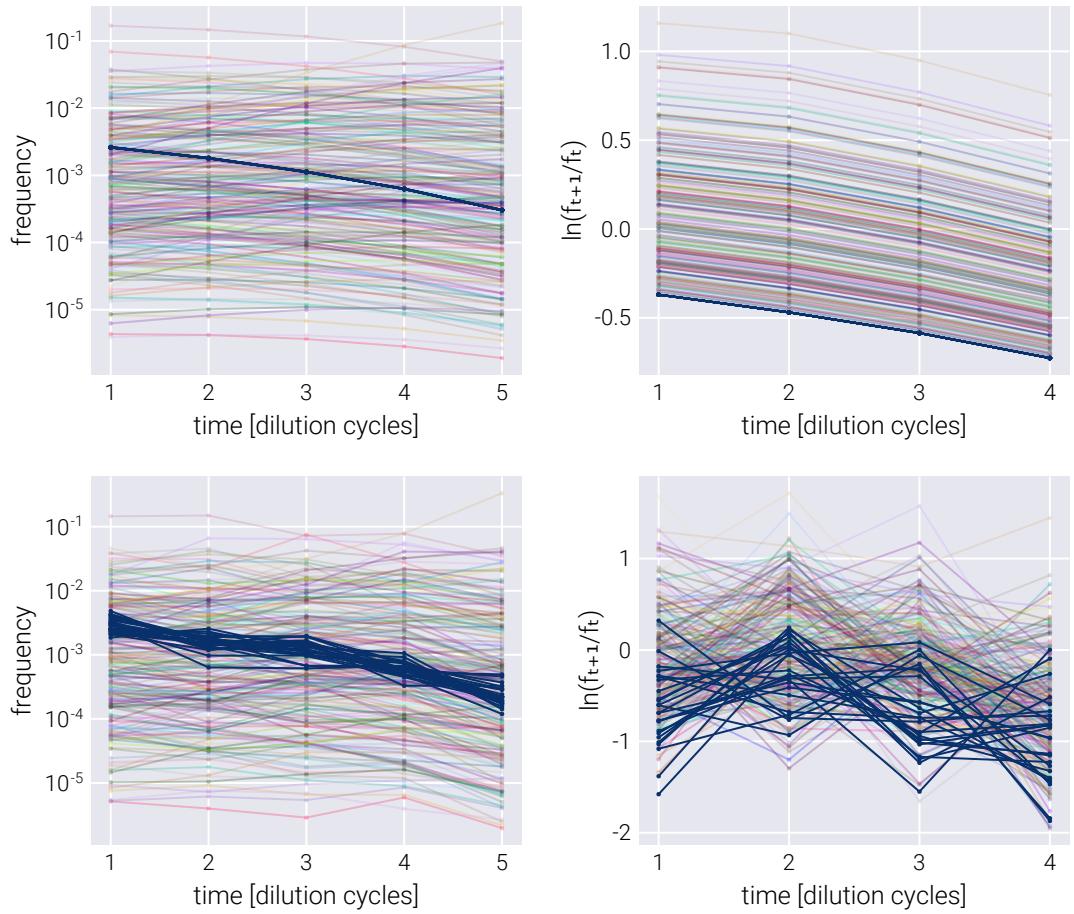


Figure S5. Logistic growth-dilution simulations with and without noise