Putting BASIL in a BLT: A Bayesian Filtering Method for Estimating the Fitness Effects of Nascent Adaptive Mutations

Huan-Yu Kuo[1,2], Sergey Kryazhimskiy[1,*]

[1]Department of Ecology, Behavior and Evolution,

[2]Department of Physics, University of California San Diego, La Jolla, CA 92093

[*]Corresponding author: skryazhi@ucsd.edu

ABSTRACT

The distribution of fitness effects (DFE) of new beneficial mutations is a key quantity that dictates the dynamics of adaptation. The barcode lineage tracking (BLT) approach is an important advance toward measuring DFEs. BLT experiments enable researchers to track the frequencies of ~10^5 of barcoded lineages in large microbial populations and detect up to thousands of nascent beneficial mutations in a single experiment. However, reliably identifying adapted lineages and estimating the fitness effects of driver mutations remains a challenge because lineage dynamics are subject to demographic and measurement noise and competition with other lineages. We show that the commonly used Levy-Blundell method for analyzing BLT data and its improved version FitMut2 can produce biased fitness estimates, particularly if selection is strong. To address this problem, we develop a new method called BASIL (BAyesian Selection Inference for Lineage tracking data), which dynamically updates the belief distribution of each lineage's fitness and size based on the number of barcode reads. We calibrate BASIL's model of noise with new experimental data and find that noise variance scales non-linearly with lineage abundance. We test how BASIL and FitMut2 perform on simulated data and on down-sampled data from the original BLT data by Levy et al and find that BASIL is both more robust and more accurate than FitMut2. Our work paves the way for a systematic inference of the distribution of fitness effects of new beneficial mutations from BLT experiments in a variety of scenarios.

AUTHOR SUMMARY

Beneficial mutations are rare but they are the ultimate drivers of evolution by natural selection. Evolutionary biologists seek to understand how many beneficial mutations an organism has access to in different environments and how these mutations affect fitness. Barcode lineage tracking (BLT) is a powerful experimental approach that tracks the frequencies of hundreds of thousands of subpopulations labeled with unique DNA barcodes and provides data that potentially enables researchers to identify and isolate many beneficial mutations arising in experimental microbial populations. However, analyzing these data is challenging because of the randomness of evolution and measurement noise. We found that existing methods for analyzing BLT data can lead to biased estimates of the fitness effects of beneficial mutations, especially when selection is strong. To overcome this issue, we developed a new method called BASIL, which uses a Bayesian approach that updates the estimated fitness and size of each lineage based on the measured barcode counts. We show that BASIL provides more accurate and robust estimates of the fitness effects of beneficial mutations in both simulated and real datasets than the existing alternatives. Thus, BASIL will facilitate a better understanding of beneficial mutations and adaptation more generally.

INTRODUCTION

Every new adaptation originates in a single individual. Most vanish soon thereafter. The few that survive may encounter fierce competition with other contending adapted lineages [1–5]. This competition—termed "clonal interference"—is particularly severe in organisms that have limited recombination and also have a large supply of beneficial mutations, such as many bacteria [6–8], cancers [9,10], and viruses [11–13]. In the clonal interference regime, the chance of fixation of a new allele depends not only on its own fitness benefit but also on the benefits provided by competing alleles [14–16]. Thus, to understand the dynamics of rapid adaptation, we must know the distribution of fitness effects (DFE) of adaptive mutations [16,17]. However, this empirical knowledge is not yet readily available.

Measuring the effects of new adaptive mutations is challenging. The most statistically sound strategy would be to isolate them from random samples of spontaneous mutations. However, in most organisms, mutations are rare [18], and beneficial mutations typically constitute a relatively small fraction of all mutations [3,19,20] (but see [21,22]). As a result, obtaining accurate statistics of beneficial mutations from samples of random mutations is generally extremely inefficient. Instead, to sample adaptive mutations, researchers leverage the power of natural selection, which elevates the frequencies of beneficial mutations in a population. It is relatively easy to find and isolate beneficial mutations from recently adapted populations, particularly those evolved in the lab [23–26]. However, this strategy preferentially captures mutations with the strongest fitness benefits because they are least likely to be lost by genetic drift or clonal interference. Therefore, this approach may provide an incomplete view of the diversity of adaptive mutations [3]. Moreover, this view may also be biased because which mutations survive and fix depends in a complex way on the DFE itself and on population size and structure [16,27,28].

The barcode lineage tracking (BLT) method is a powerful strategy for estimating the fitness effects (also referred to as selection coefficients) of many nascent beneficial mutations in asexual genetically tractable microbes [3–5,29–36]. The key idea underlying BLT is to introduce many (typically ~10^5) random neutral DNA barcodes into an otherwise isogenic population and track barcode frequencies over time using deep sequencing [3,37]. Each new adaptive mutation that arises in the population is permanently linked to a unique DNA barcode. As an adaptive mutation spreads in the population, the corresponding increase in the frequency of the linked barcode can be detected in the sequencing data. Moreover, since the sizes of most barcoded subpopulations are initially small (typically ~100 individuals), the expansion of each adapted subpopulation is initially driven by a single adaptive mutation. In other words, for a period of time, each barcode with increasing frequency reports on a single adaptive driver mutation. As a result, this approach allows one to measure the fitness benefits of many simultaneously segregating driver mutations, including weak ones, in a single relatively short evolution experiment, while avoiding laborious genetic reconstructions and screens [3,38]. Although the BLT approach does not entirely eliminate biases introduced by natural selection, it significantly mitigates them, and overall provides a much more accurate picture of the diversity of adaptive mutations than other existing methods [3].

While the BLT approach significantly advances our ability to capture and study adaptive mutations, analyzing BLT data remains difficult. The main challenge is to distinguish lineages that have acquired a single beneficial mutation ("adapted lineages") from those that have not ("neutral lineages") based on noisy dynamics of low-abundance barcodes. The core idea is the same as in classical competition assays [39–42], i.e., to determine whether a lineage systematically increases in frequency relative to neutral reference lineages (i.e., those with wildtype fitness) and then to infer

the fitness effect of the underlying driver mutation from the rate of this increase. However, despite many similarities, the BLT setup differs in three important ways from the canonical competition assay. First, in competition assays, lineages are usually represented by thousands of cells at the beginning of the experiment, whereas in the BLT experiments all lineages are by design present at low abundances (~100 cells), which makes lineage extinctions much more likely. Second, competition assays are typically short ($\lesssim$ 50 generations) to prevent new mutations from significantly affecting lineage dynamics, whereas BLT experiments are longer (> 100 generations) precisely to allow for new adaptive mutations to arise and reach sufficiently high frequencies. Third, in competition assays, designated neutral reference lineages are deliberately added to the population. Their frequency dynamics reports on population's mean fitness which allows one to estimate the fitness of all other lineages relative to the reference [22,41]. In contrast, in the BLT setup, all lineages are initially equivalent, and since adaptive mutations can arise during the experiment (or shortly before), which lineages become adapted and which ones remain neutral is a priori unknown.

The absence of defined neutral reference lineages is perhaps the biggest challenge of analyzing BLT data because such lineages provide the most obvious and robust way to readout the mean fitness of the population, which in turn is required for the inference of fitness of all other lineages. Two main approaches addressing this problem have been proposed so far. Levy et al developed the original approach, which we refer to as the "Levy-Blundell" or the "LB" method for short [3]. They chose low-abundance lineages as the neutral reference, reasoning that such lineages are unlikely to acquire adaptive mutations during the BLT experiment, and inferred mean fitness based on the rate of decline in the frequency of such lineages. However, low-abundance lineages are likely to go extinct, especially if selection is strong. One could choose more abundant lineages as the neutral reference, but such lineages are more likely to carry adaptive mutations, especially later in the experiment. Both outcomes are undesirable as they can bias the estimates of fitness effects of all adaptive mutations. More recently, Li et al developed an approach called FitMut2, which retains many features of the original LB method but instead of relying on designated neutral reference lineages it iteratively switches between identifying adapted lineages and estimating mean fitness on their basis until the process converges [43]. They have shown that FitMut2 outperforms the original LB approach on their simulated data. However, FitMut2 could also be prone to biases because the solution to which it converges might be sensitive to initial conditions and/or noise in the data.

Given these potential theoretical concerns, it is important to assess the performance of existing approaches empirically on simulated data where the ground truth is known, and to understand when and why they fail. In particular, the ability to accurately infer fitness effects will likely depend on the strength of selection because stronger selection accelerates evolutionary dynamics and reduces the amount time available for sampling lineage trajectories. To this end, in the first part of this paper, we assess the performance of the LB and FitMut2 approaches on BLT data simulated under weak and strong selection. We find that both approaches perform well under weak selection but produce substantially biased fitness estimates under strong selection. Since the LB method has been more widely used [3,4,29,31,32,35], we carry out a deeper investigation of the underlying reasons for its poor performance.

In the latter part of the paper, we develop a new BLT analysis method termed BASIL (BAyesian Selection Inference for Lineage tracking data), which keeps track of belief distributions for the fitness of all lineages under the assumption that most lineages are initially neutral. It updates these distributions based on barcode read counts, and uses random lineages (both putatively adapted and

neutral) to estimate mean fitness. Our approach combines several techniques developed previously for the analyses of competition assays [42,44–46] and BLT experiments [3,5,43]. In particular, similar to previous methods, BASIL models the evolutionary dynamics of lineages between sampling time points as well as the measurement process, whose properties we measure using a new calibration experiment. In contrast to most other approaches and similar to Ref. [45], BASIL treats unobserved lineage sizes as hidden variables, which is important for obtaining accurate estimates of lineage fitness, particularly under strong selection. Similar to Ref. [42], we cast our model in the Bayesian framework, which allows us to keep track of uncertainties in our fitness and lineage size estimates and dynamically update our belief distributions as the data arrives. We then directly compare the performance of BASIL on published BLT datasets to the performance of FitMut2. Finally, we use BASIL to characterize adaptation of several strains of yeast *Saccharomyces cerevisie* to various environments.

RESULTS

Existing BLT analysis methods can produce biased fitness estimates

To understand how accurately the LB and FitMut2 methods infer the fitness effects of nascent adaptive mutations, we simulate a batch-culture BLT experiment with 10^5 barcoded lineages, mimicking the setup of Levy et al [3]. The details of our simulations are provided in Materials and Methods. Briefly, each batch-culture cycle begins with dilution, which we simulate by down-sampling each lineage to $1/D = 1/256$ of its size using the Poisson distribution. Some lineages may go extinct at this step. Then, all surviving lineages deterministically expand by a factor proportional to the difference between their fitness and the population's mean fitness, so that the population reaches approximately its pre-dilution size. Our simulations continue for 20 cycles, corresponding to 160 generations. Every other cycle, we simulate barcode sequencing by randomly sampling a certain number of individuals from our population and recording their lineage identities.

In our simulations, 3,000 lineages are adapted, i.e., all individuals within these lineages have the same fitness advantage compared to the rest of the population. This scenario where all beneficial mutations are pre-existing is less complex than many real BLT experiments, and inference methods should perform best in this case. Our reasoning for using these relatively simple simulations was twofold. First, many if not most adaptive mutations detected in BLT experiments arise prior to the beginning of the experiment [3,4,33]. Second, we reasoned that if we find that a method fails on these relatively simple simulated data, it is very unlikely to perform well on real BLT datasets. Since there are no new mutations, our simulations are similar to competition assays, with the exception that the identity of neutral and adapted lineages is unknown to the inference algorithm. We simulate two conditions that we refer to as "strong selection" and "weak selection" (Figure S1 and Table S3). In the weak selection regime, which resembles the original BLT experiment carried out by Levy et al [3], the average fitness of an adapted lineage is 3%. As a result, the mean fitness of the population increases by about 4% in 150 generations. In the strong selection regime, the average fitness of an adapted lineage is 8%, and the population's mean fitness increases by about 12% in 150 generations.

Our first goal is to assess how the LB and FitMut2 methods perform on these simulated data. If the methods underperform, we would like to understand why, focusing specifically on the better-established LB method. To enable a potential in-depth investigation, we simplify the original LB approach by stripping away much of the complexity required for analyzing real BLT datasets but

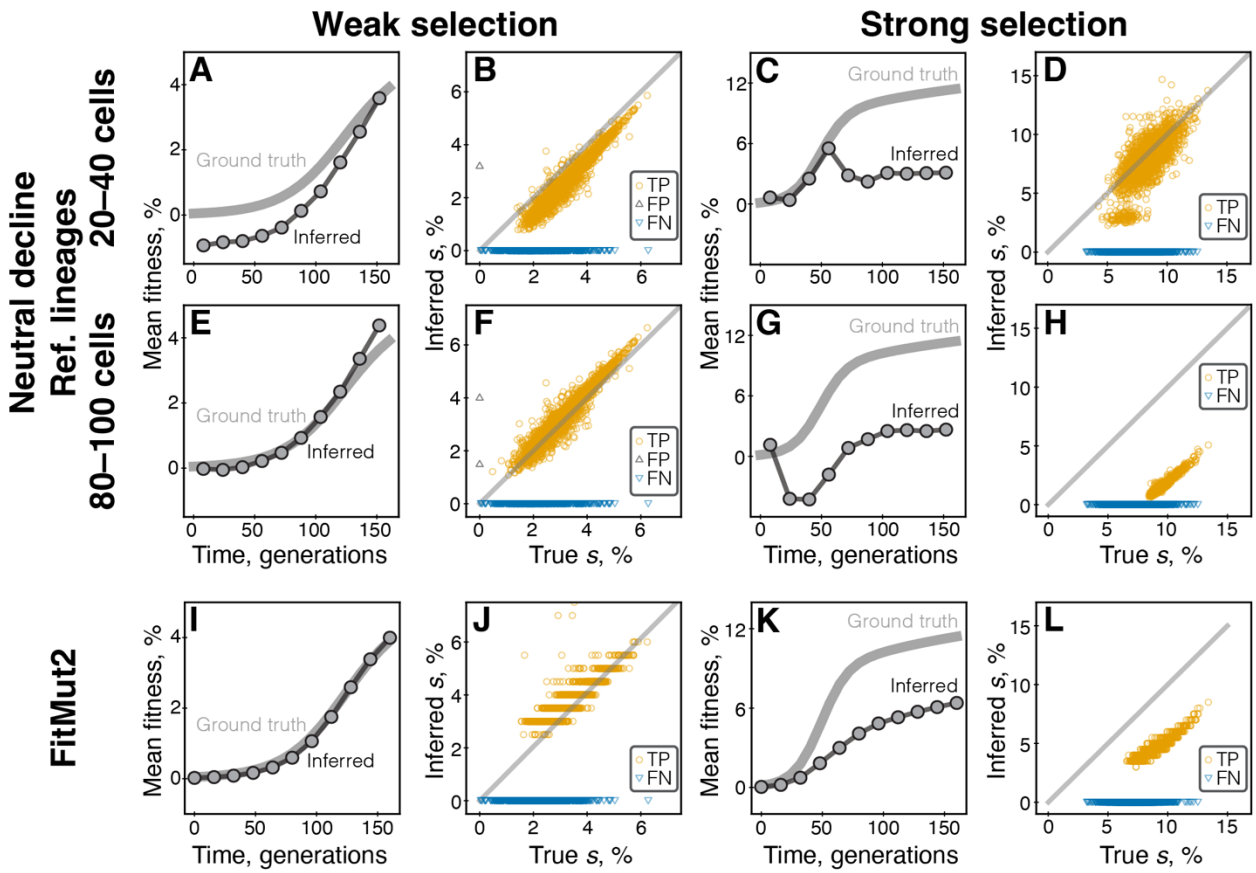

Figure 1. Performance of the neutral decline and FitMut2 methods on simulated BLT data depends on the strength of selection. Panels A–H show the results for the neutral decline method, and panels I–L show the FitMut2 results. Left panels A, B, E, F, I, J show simulations in the weak selection regime; right panels C, D, G, H, K, L show simulations in the strong selection regime. **A, C.** Mean fitness trajectories, true and inferred by the neutral decline method with low-abundance reference lineages. **B, D.** Selection coefficients of adapted lineages, true versus inferred by the neutral decline method with low-abundance reference lineages. Each symbol corresponds to a lineage; orange circles, black upward triangles and blue downward triangles represent true positives, false positives and false negatives, respectively; true negatives are not shown. **E–H.** Same as A–D but for high-abundance lineages. **I–L.** Same as A–D but for FitMut2.

unnecessary for the analysis of our simulated data, while retaining its core assumptions (see Materials and Methods and Section 2.3 of the Supplementary Information). We refer to this simplified version as the "neutral decline" method to distinguish it from the original. As the original LB approach, the neutral decline method requires us to specify which lineages we select as the neutral reference. To determine how this choice might affect inference, we select either low- or high-abundance lineages, with low-abundance lineages being defined as those represented by 20 to 40 reads (corresponding to 20 to 40 cells at the bottleneck in our simulations) and high-abundance lineages being defined as those represented by 80–100 reads (corresponding to 80 to 100 cells at the bottleneck in our simulations).

The performance of both the neutral decline and FitMut2 methods on our BLT simulations are illustrated in Figure 1. We find that in the weak selection regime, both FitMut2 and the neutral decline method with high-abundance reference lineages accurately infer the mean fitness trajectory

and the fitness effects of individual adaptive mutations (Figure 1E,F,I). When low-abundance lineages are used as reference, the neutral decline approach underestimates both the mean fitness of the population and the selection coefficients of individual adaptive mutations even in this favorable regime (Figure 1A,B). Both methods perform significantly worse in the strong selection regime (Figure 1C,D,G,H,K,L). FitMut2 and the neutral decline approach with high-abundance reference lineages severely underestimate both the mean fitness and the effects of individual mutations (Figure 1G,H). If low-abundance lineages are used as reference, the neutral decline method estimates the mean fitness accurately, but only for the first ~50 generations (Figure 1C), which results in noisy estimates of selection coefficients of individual adaptive mutations (Figure 1D).

Thus, the performance of existing BLT analysis methods depends on the conditions of the BLT experiment, in particular on the strength of selection. Furthermore, the accuracy of the neutral decline method—and by extension that of the LB method—hinges on the choice of reference lineages.

Statistical causes for bias of the LB method

To understand why the neutral decline method (and, consequently, the LB method) sometimes produces biased estimates of lineage fitness, we first summarize the central assumption underlying this approach (see Materials and Methods and Table S4 for more details). To infer the population mean fitness $\bar{s}_k$ at the sampling interval (t_{k-1}, t_k), the neutral decline method groups together reference lineages with each reads count r_{k-1} observed at t_{k-1}. For each such group, it computes the average read count $\bar{r}_k$ at the next sampling time t_k and infers mean fitness $\bar{s}_k$ from the equation

$$\frac{\bar{r}_k}{R_k} = \frac{r_{k-1}}{R_{k-1}} e^{-\bar{s}_k \Delta t_k}, \tag{1}$$

where R_k is the total read depth at time t_k and $\Delta t_k = t_k - t_{k-1}$. Equation (1) can be derived from the standard population genetics theory under the crucial assumption that all lineages with the same number of reads r_{k-1} are present in the population at frequency r_{k-1}/R_{k-1} at t_{k-1} [47]. We refer to equation (1) as the "neutral decline" equation. Once the mean fitness trajectory $\bar{\boldsymbol{s}} = (\bar{s}_1, \bar{s}_2, \dots)$ is known, the neutral decline method infers the selection coefficients of all non-reference lineages by maximizing the likelihood of their frequency trajectories (for details, see Materials and Methods). Importantly, even though the original LB method is more sophisticated in that it utilizes the entire conditional distribution of lineage read counts at the next sampling time given their read counts at the previous time point (see equation (45) in the Supplementary material to Ref. [3] or equation (S8) in the Supplementary Information), the mean of that distribution is given by equation (1). Therefore, a failure of equation (1) to capture the relationship between r_{k-1} and $\bar{r}_k$ would imply a failure of the original LB model.

To test the validity of equation (1), we plot the rescaled average read count $\tilde{r}_k = \bar{r}_k R_{k-1}/R_k$ at time point t_k against the corresponding observed read count r_{k-1} at the previous time point. According to equation (1), $\tilde{r}_k$ must be proportional to r_{k-1} with a zero y-intercept. Instead, we find that this relationship is more complex, with its shape being dependent on the selection regime and on time (Figure 2A,C). Specifically, when selection is weak, $\tilde{r}_k$ relates linearly to r_{k-1} during the entire BLT experiment, but with a non-zero y-intercept (Figure 2A). Despite its relatively small value (e.g., 6.60 at generation 16 and 2.89 at generation 160), a non-zero y-intercept can cause significant underestimates of mean fitness when low-abundance lineages are chosen as reference (Figure 2B), including negative estimates at early time points (see Figure 1A). Since the estimates of selection coefficients of all lineages depend on the estimate of mean fitness, the bias in the latter propagates

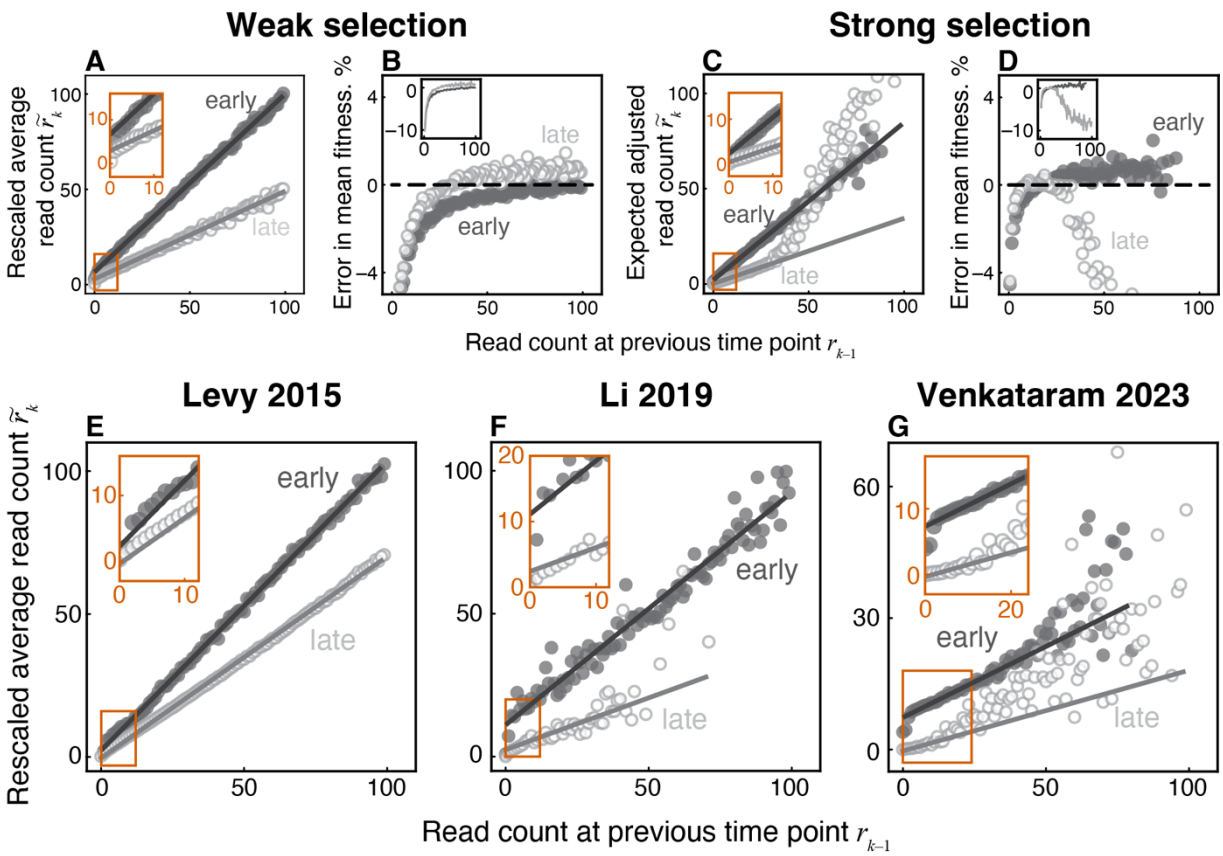

Figure 2. Observed lineage read counts deviate from those predicted by equation (1) both in simulated and real data. A. Rescaled average read count $\tilde{r}_k = \tilde{r}_k R_{k-1}/R_k$ at the next time point (ordinate) plotted against the read count observed at the previous time point r_{k-1} (abscissa) for the weak-selection simulation. Light and dark gray points represent early ($t_k = 16$ generations) and late ($t_k = 160$ generations) time intervals. Lines represent least-squares best fits within the linear regime (see Materials and Methods). **B.** Error in the inferred mean fitness (inferred minus true) as a function of the read count of lineages that are used as the neutral reference. Shades are the same as in panel A. **C.** Same as panel A but for the strong selection regime. Early time interval is at $t_k = 16$ generations and late time interval is at $t_k = 64$ generations. **D.** Same as panel B but for the strong selection regime. **E–G.** Same as panel A but for three real BLT datasets: Levy 2015 R1 (panel E), Li 2019 Evo1D R2 (panel F) and Venkataram 2023 Co-evolution R5 (panel G). Early time interval is at $t_k = 16, 7, 20$ generations in panels E, F, G, respectively and late time interval is at $t_k = 112, 133, 86$ generations.

to a bias in the former, as seen in Figure 1B. Choosing initially more abundant lineages as reference produces a more accurate estimate of initial mean fitness (Figures 1E and 2B) because such estimates are less sensitive to the value of the y-intercept (Figure 2A). However, choosing more abundant lineages is not universally better. For example, under strong selection, the relationship between $\tilde{r}_k$ and r_{k-1} becomes non-linear for higher-abundance lineages, particularly later in the experiment (Figure 2C), which also biases the estimates of mean fitness (Figures 1C,G and 2D).

These observations show that the relationship between lineage read counts at successive time points can be used to diagnose potential inference problems in the absence of the ground truth. Thus, we examined these relationships in real BLT data, asking whether they exhibit similar deviations from equation (1) as we observed in our simulations. To this end, we reanalyzed data from three published BLT studies, Levy 2015 [3], Li 2019 [29] and Venkataram 2023 [33]. We find

strong deviations from linearity in the Li 2019 and Venkataram 2023 data for higher-abundance lineages at later time points (Figures 2 F,G). Even at the initial time point where we expect equation (1) to be most accurate, we find that, even though $\tilde{r}_k$ depends on r_{k-1} linearly, there is a statistically significant positive y-intercept (2.26 for Levy 2015, 11.10 for Li 2019, and 7.32 for Venkataram 2023, all P-values $< 10^{-12}$, F-test; Figure 2E–G). As a result, the estimates of mean fitness and selection coefficients of adaptive mutations in real data produced by the LB method are likely subject to the same biases as observed in simulated data sets.

Finally, we asked why the observed relationship between r_{k-1} and $\bar{r}_k$ deviates from the prediction given by equation (1). First, the deviation from linearity at high r_{k-1}, especially later in the experiment (Figure 2C,F,G), can be explained if reference lineages are in fact not neutral. Indeed, after selection had time to act, lineages that carry stronger beneficial mutations are typically present at higher frequencies in the population and hence have higher values of r_{k-1} than neutral lineages. As a result, lineages with higher r_{k-1} increase disproportionately more or decrease disproportionately less than lineages with lower r_{k-1}. Second, to understand why the relationship between r_{k-1} and $\bar{r}_k$ has a non-zero y-intercept, consider a hypothetical example where all lineages are neutral and equally abundant in the population and the coverage at both time points t_{k-1} and t_k is such that a typical lineage is represented by 10 reads. Then, at t_{k-1}, while most lineages are represented by $r_{k-1} = 10$ reads, measurement noise will result in many lineages with $r_{k-1} = 9$, $r_{k-1} = 11$, $r_{k-1} = 8$, $r_{k-1} = 12$, etc. Since all lineages are neutral, the expected read count for any lineage at the next time point is $\bar{r}_k = 10$, including those with $r_{k-1} \neq 10$, whereas the neutral decline equation (1) predicts $\bar{r}_k = r_{k-1}$. This bias arises as a result of the regression to the mean, whereby lineages whose read counts are by chance abnormally low at one time point are represented by a more typical (larger) number of reads at the next time point, causing the y-intercept to be non-zero (see Supplementary Information, Section 2.3.2 for an extended discussion). The specific value of the intercept depends, among other things, on the measurement noise and on the distribution of lineage frequencies in the population, quantities that are a priori unknown. Therefore, correcting for this intercept appears difficult.

Overall, this investigation demonstrates that the neutral decline approach (and, consequently, the LB method) can lead to biased inferences of the fitness effects of nascent adaptive mutations because equation (1) (or the corresponding equation (S8) for the full distribution of reads) is in general incorrect, particularly for low-abundance lineages or under strong selection.

BASIL: Bayesian selection inference for lineage tracking data

To overcome challenges of the existing approaches, we developed BASIL (Bayesian Selection Inference for Lineage tracking data), a robust statistical method for identifying adapted lineages and inferring their fitness and implemented it in a software package [48]. The key hidden variables in BASIL are the sizes n and selection coefficients s (fitness) of individual lineages. Population's mean fitness $\bar{s}_k$ and the measurement noise parameter ϵ_k defined below are time-dependent global variables. At any given sampling time t_k, each lineage i is characterized by the joint belief distribution $P_{ik}^{\text{belief}}(n, s)$ for its size and fitness, which we model parametrically as a product of a normal and gamma distributions (see Materials and Methods and equations (S31)–(S33) in the Supplementary Information). BASIL is a Bayesian filtering model (Figure 3B, [49]), in which the belief distributions $P_{ik}^{\text{belief}}(n, s)$ are based on the past data, i.e., data accumulated up to and including time t_k.

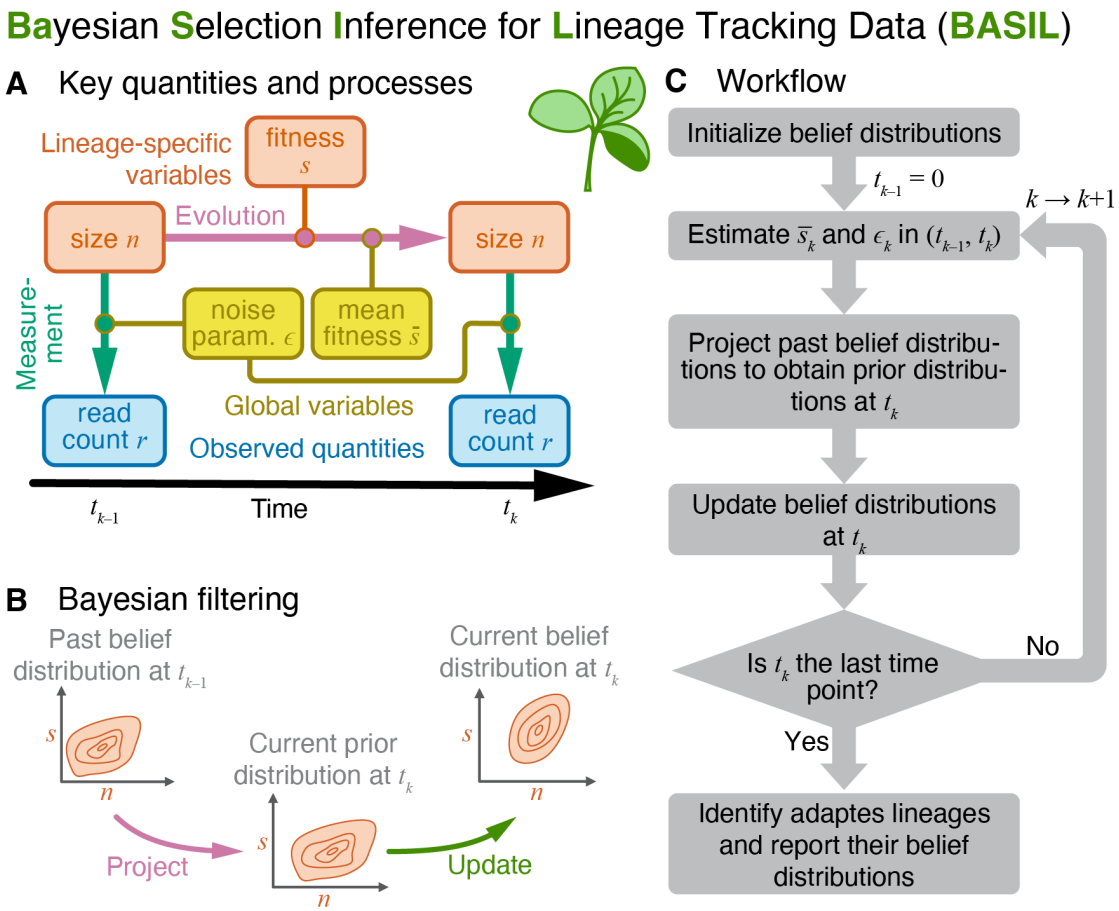

Figure 3. BASIL schematic. A. Key quantities and processes. Each barcoded lineage i is characterized at the current time t_k by its size n_{ik} and fitness (selection coefficient) s_{ik}, which are unknown. Lineage size changes over time depending on the lineage fitness and the mean fitness of the population. At each sampling time point, we measure relative lineage sizes by counting sequencing reads with the corresponding barcode. **B.** Bayesian filtering consists of projecting the past belief distribution over the hidden variables n_{k-1} and s_{k-1} using the model of evolution to obtain the prior distribution for n_k and s_k for the current sampling time point t_k and then updating this distribution using read count r_k to obtain the new belief distribution. **C.** BASIL workflow.

At BASIL's core is the model of evolutionary dynamics that computes the prior distribution $P_{ik}^{\text{prior}}(n, s)$ for the size and selection coefficient of each lineage i at the current sampling time t_k based on the past belief distribution $P_{ik-1}^{\text{belief}}(n, s)$ and the population's mean fitness $\bar{s}_k$ in the time interval (t_{k-1}, t_k) (Figure 3A). We also develop a model of measurement $P^{\text{meas}}(r|n; \epsilon)$, which probabilistically relates the true lineage size n to the corresponding barcode read count r observed in the sequencing data. We model $P^{\text{meas}}(r|n; \epsilon)$ as a negative binomial distribution that depends on the current value of the noise parameter ϵ.

Estimation of $P_{ik}^{\text{belief}}(n, s)$ at t_k occurs in three steps (Figure 3C). First, we estimate the mean fitness of the population $\bar{s}_k$ and the noise parameter ϵ_k in the current time interval (t_{k-1}, t_k) by sampling thousands of random lineages, computing their prior distributions $P_{ik}^{\text{prior}}(n, s)$ and then maximizing the likelihood of their observed read counts r_{ik} at t_k. Second, we compute the prior distributions

$P_{ik}^{\text{prior}}(n,s)$ for all lineages, given the estimated population's mean fitness $\bar{s}_k$. We then apply the Bayes theorem to obtain the updated lineage belief distributions $P_{ik}^{\text{belief}}(n,s)$, given the read counts r_{ik} and the estimated noise parameter ϵ_k. In the rest of the section, we flesh out the main ideas and expressions behind our method. The full mathematical details are provided in Section 3 of the Supplementary Information.

Model of evolution. Consider a barcoded lineage with fitness s relative to the ancestor and suppose that this lineage is represented by $n_{k-1} = n(t_{k-1})$ individuals in the population at the previous sampling time t_{k-1}. On average, such lineage will expand if s is larger than the population's current mean $\bar{s}$ and it will contract if $s < \bar{s}$. Demographic noise will also cause its size to randomly fluctuate around this expectation, such that the lineage size n_k at the current sampling time t_k is a random variable. To model it, we assume that our BLT population grows in the batch-culture regime [25] and sampling happens immediately prior to a dilution step. Then, during dilution, a fraction $1/D$ of the population is transferred into the fresh medium and the rest is discarded, where $D > 1$ is the dilution factor. Immediately after dilution, the size of the focal lineage becomes $\tilde{n}$, which we draw from the Poisson distribution with mean n_{k-1}/D. Then, we assume that the lineage grows or shrinks exponentially and deterministically for the batch-culture cycle duration Δt_c, such that by the end of the cycle its size becomes $n(t_{k-1} + \Delta t_c) = \tilde{n} D e^{(s-\bar{s})\Delta t_c}$. If the next sampling point is the end of this cycle, then $n_k = n(t_{k-1} + \Delta t_c)$. Otherwise, we repeat the dilution and growth phases until the sampling time t_k is reached. Note that this model allows for the possibility of lineage extinction during dilution.

Model of measurement. The actual lineage sizes are not observable. Instead, the population is sampled at time points t_0, t_1, … and sequenced at the barcode locus, which involves a series of processing steps [50]. Then the sequenced reads containing each barcode are counted. This measurement process can be described by a probability distribution $P^{\text{meas}}(r|n)$ of observing r reads for a barcode lineage with n cells. The simplest model for this process is the Poisson distribution with mean nR/N, where R is the read depth (i.e., the total number of reads obtained for the sample) and N is the total population size at the sampling point. However, sequencing read counts are often overdispersed with respect to the Poisson distribution [46,51]. Previously, Levy et al modeled measurement noise with a distribution that allows for overdispersion and in which variance scales linearly with the mean, similar to the Poisson distribution [3]. Other studies modeled read counts with a negative binomial distribution which permits an arbitrary scaling between mean and variance [52–54].

To determine how measurement noise variance scales with the mean read count, we performed the following calibration experiment. We assembled a barcoded population of *Saccharomyces cerevisiae* with a total size of 1.6×10^7 individuals that consisted of 26 subpopulations present at six different frequencies (from 10^{-5} to 0.40) and sequenced it with 9-fold replication to an average depth of about 2×10^5 reads per replicate (see Materials and Methods for details).

As expected, we found that the mean of the barcode frequency estimated from sequencing data is very close to the input frequency of each lineage (Figure 4A). To test the extent of overdispersion, we plotted the variance of the measured read counts σ_r^2 against the mean $\langle r \rangle$. If the measurement process was adequately described by the Poisson distribution, we would expect to observe a linear relationship with slope 1, whereas the noise model of Levy et al predicts linear scaling with a slope

greater than 1 [3]. We find that the Poisson model fits our data quite well for read counts $\lesssim$100 but fails for more abundant barcodes (Figure 4B). Instead, the variance is better fit by a non-linear convex function of the mean. We capture this nonlinearity with the model

$$\sigma_r^2 = \langle r \rangle + \epsilon \langle r \rangle^2, \qquad (2)$$

where ϵ is a measurement noise parameter that controls the degree of overdispersion (Figure 4B, blue line). We estimate that $\epsilon = 2.4\times10^{-3}$ in the calibration experiment, which is significantly different from zero (Table S2). Since the specific value of ϵ might depend on experimental details, we consider equation (2) as a constraint on our model of noise, and allow ϵ to be a free parameter at each time interval. Since the functional form (2) requires a more flexible distribution than afforded by either the Poisson or the Levy et al distributions, we model the measurement process with a negative binomial (NB) distribution (equation (5) in Materials and Methods) with mean $\langle r \rangle = nR/N$ and the variance given by equation (2).

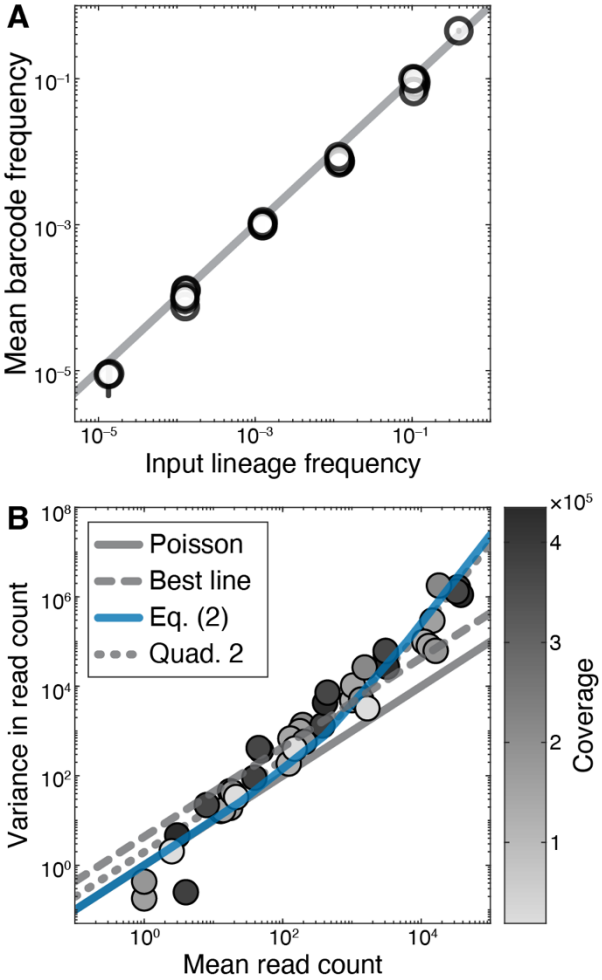

Figure 4. Properties of the measurement process. A. Barcode frequency estimated from barcode sequencing data (abscissa) versus the lineage frequency set experimentally (ordinate). Each point is a unique barcode. Error bars show one standard deviation of the mean. The 1-to-1 line is shown in gray. **B.** Variance of the read count versus the mean read count. Each point represents the data from multiple barcodes of the same lineage frequency. Shade represents coverage. Solid blue line is the parabola given by equation (2). Other curves are described in the Materials and Methods and their fitted parameter values are provided in Table S2.

Bayesian filtering. For each lineage I, we seek to estimate the joint belief distributions $P_{ik}^{\text{belief}}(n,s)$ of lineage size n and fitness effect s at each sampling time t_k. First, assuming that the past belief distribution $P_{ik-1}^{\text{belief}}(n,s)$ at t_{k-1} and the mean fitness $\bar{s}_k$ in the time interval (t_{k-1}, t_k) are known, we obtain the prior distribution $P_{ik}^{\text{prior}}(n,s;\bar{s}_k)$ at the current time t_k by projecting the past belief distribution P_{ik-1}^{belief} using our model of evolution. The prior distribution P_{ik}^{prior} represents our best guess for the abundance of lineage I and its fitness at the current time point t_k, based on its past trajectory and the population's mean fitness. The observed read count r_{ik} for this lineage may or may not be consistent with this guess, and we use this new data to update our belief distribution using the Bayes' theorem (Figure 3B),

$$P_{ik}^{\text{belief}}(n,s) = \frac{1}{Z_{ik}^{\text{sel}}} P^{\text{meas}}(r_{ik}|n;\epsilon_k) P_{ik}^{\text{prior}}(n,s;\bar{s}_k) \text{ for k} \geq 1, \qquad (3)$$

where P^{meas} is the NB distribution defined in the "Model of measurement" section and Z_{ik}^{sel} is the marginal probability that r_{ik} reads are observed at t_k, given all of our prior knowledge (subscript S

stands for "selection" and refers to the fact that this model allows for the fitness of the lineage to be non-zero). Note that Z_{ik}^{sel} depends on the noise parameter ϵ_k and the mean fitness $\bar{s}_k$ in the current time interval (t_{k-1}, t_k), which are assumed to be known.

Initialization. We initialize the belief distribution for any lineage i as $P_{i0}^{\text{belief}}(n,s) = \gamma_i(n)N(s)$, where $\gamma_i(n)$ is a Gamma distribution with parameters based on the lineage's initial read count and $N(s)$ is the normal distribution with zero mean and standard deviation 0.1. In other words, we assume that all lineages are initially most likely neutral, but with large uncertainty.

Estimating population's mean fitness. As mentioned above, our procedure for updating the belief distribution at time t_k depends on the mean population fitness $\bar{s}_k$ and the noise parameter ϵ_k at the time interval (t_{k-1}, t_k). To estimate $\bar{s}_k$ and ϵ_k, we randomly choose 3,000 lineages and classify them as either putatively adapted (set A_k) or putatively neutral (set N_k). A lineage is classified as putatively adapted if the mean of its marginal belief distribution $P_{ik-1}^{\text{belief}}(s)$ is positive and at least three times larger than its standard deviation; and it is classified as putatively neutral otherwise. We then estimate $\bar{s}_k$ and ϵ_k by maximizing the log-likelihood function

$$L_k(\bar{s},\epsilon) = \sum_{i\in A_k} \log Z_{ik}^{\text{sel}}(\bar{s},\epsilon) + \sum_{i\in N_k} \log Z_{ik}^{\text{neut}}(\bar{s},\epsilon), \tag{4}$$

where Z_{ik}^{neut} is a quantity analogous to Z_{ik}^{sel} but for the neutral model where the selection coefficient of the lineage is zero.

Identification of adapted lineages. As a result of the procedure described above, for each lineage i, we obtain a time-varying belief distribution $P_{ik}^{\text{belief}}(n,s)$, $k = 1, 2, \ldots$ We calculate the final marginal belief distribution for the selection coefficient of lineage i as $P_i^{\text{belief}}(s) = \int_0^\infty P_{ik'(i)}^{\text{belief}}(n,s)dn$ choosing the time point $t_{k'(i)}$ where the s-variance of the belief distribution $P_{ik}^{\text{belief}}(n,s)$ is minimal. The choice to minimize s-variance (rather than simply using the last available time point) is motivated by the fact that secondary adaptive mutations arising later in the experiment may cause widening of the belief distribution. While such widening may be biologically informative, the main purpose of BLT experiments is to estimate the effects of single nascent adaptive mutations, suggesting that one should select a time point before secondary mutations become sufficiently abundant. In practice, this choice appears unimportant, as the s-variance is minimized at the last time point for the overwhelming majority of lineages across all datasets that we analyzed (see Data S1, Tab 3).

Once we obtain the lineage's final belief distribution $P_i^{\text{belief}}(s)$, we call lineage i adapted if the mean $\hat{s}_i$ of this distribution is sufficiently separated from 0, that is, if $\hat{s}_i \geq \beta\sigma_i$, where σ_i is the standard deviation of $P_i^{\text{belief}}(s)$. We refer to β as the "confidence factor", a hyper-parameter that must be determined empirically. Intuitively, larger β would require a lineage to have a higher fitness to be recognized as adapted, which would increase precision but reduce recall, or, equivalently, reduce the rate of false positives at the expense of increasing the rate of false negatives. In other words, a higher confidence factor would allow us to be more confident that the lineages identified as adapted are in fact adapted, albeit at the expense of missing more adapted lineages for which evidence of adaptation is weaker. On the other hand, decreasing β would increase recall but reduce precision, or, equivalently, it would allow us to identify more adapted lineages, albeit at a cost of also misidentifying more neutral lineages as adapted. Recall and precision usually exhibit a trade-off, and there is no general principle for choosing β. In the next section, we test our method on simulated data, which provides us with a guideline for the value of the confidence factor.

Comparison of methods on simulated data

For an initial assessment of BASIL performance and to determine the value of the confidence factor β, we applied BASIL to simulated BLT data described above (see Section "Estimation of mean fitness from the decline of neutral lineages can produce biased estimates of fitness effects" and Section 4 of the Supplementary Information). We found that BASIL accurately captures the dynamics of mean fitness in both regimes (compare Figures 1 and 5), even at relatively late time points where the other approaches fail. To determine the hyper-parameter β, we varied it between 0.0 and 6.4 with the step size of 0.1 and calculated the F1 score for adapted lineage calls, which is the harmonic mean of recall and precision (Figure 5E,G). We found that the F1 score is maximized at $\beta = 3.4$ and $\beta = 3.2$ for the weak and strong selection regimes, respectively. Thus, we chose the value of $\beta = 3.3$ for all our subsequent analyses (see Section 4.1 in the Supplementary Information and Figure S5). Using this confidence factor, we identified 2,570 adapted lineages (430 false negatives) in the weak selection regime, with 31 false positives, resulting in 98.8% precision and 85.7% recall. In the strong selection regime, we identified 2,481 adapted lineages (519 false negatives) with 7 false positives, resulting in 99.7% precision and 82.7% recall. For lineages identified

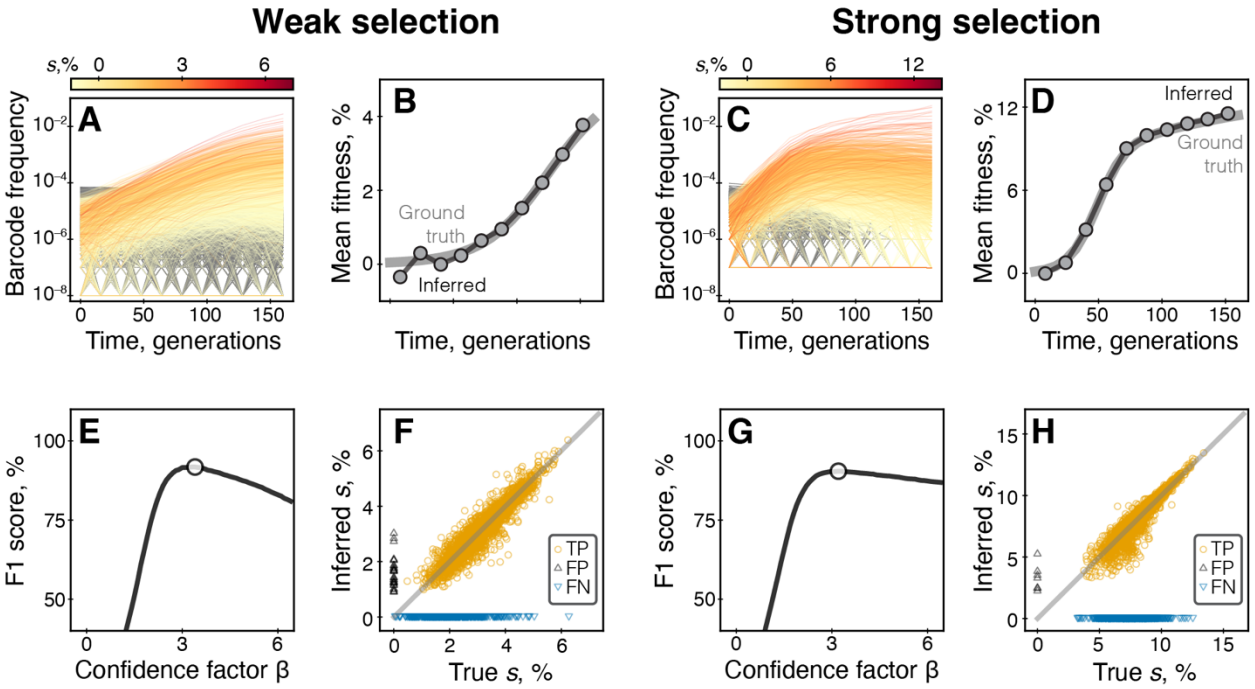

Figure 5. Performance of BASIL on simulated data. Left panels A, B, E, F show simulations in the weak selection regime; right panels C, D, G, H show simulations in the strong selection regime. **A, C.** Observed barcode frequency trajectories colored by fitness. **B, D.** Mean fitness trajectories. Grey lines show the ground truth, circles show values inferred by BASIL. **E, G.** F1 score as a function of the confidence factor β. White circle denotes the maximal value. **F, H.** Inferred versus true selection coefficients of individual lineages. Lineages are called adapted with $\beta = 3.3$ (see main text). Each symbol corresponds to a lineage; orange circles, black upward triangles and blue downward triangles represent true positives, false positives and false negatives, respectively; true negatives are not shown.

Measure of performance	Weak selection				Strong selection			
	Neutral Decline		Fit-Mut2	BASIL	Neutral Decline		Fit-Mut2	BASIL
	LA Ref[1]	HA Ref[2]		β = 3.3	LA Ref[1]	HA Ref[2]		β = 3.3
TP[3]	2204	2426	1871	2570*	1189	643	1279	2481*
FP[4]	37	49	19*	31	1	1	0*	7
FN[5]	796	574	1129	430*	1111	2357	1721	519*
Precision[6], %	98.3	98.0	99.0*	98.8	99.9	99.8	100.0*	99.7
Recall[7], %	73.5	80.9	62.4	85.7*	63.0	21.4	42.6	82.7*
F1 score[8], %	84.1	88.6	76.5	91.8*	77.3	35.2	59.8	90.4*
Error in $\bar{s}$[9], %	0.75	0.19	0.07*	0.12	4.69	7.07	3.65	0.16*
Error in $\hat{s}$[10], %	0.64	0.27	0.40	0.24*	1.29	7.99	4.41	0.39*

Table 1. Comparison of BLT analysis methods on simulated data. [1]Low-abundance (LA) lineages used as reference; [2]High-abundance (HA) lineages used as reference; [3]Number of true positives; [4]Number of false positives; [5]Number of true positives; [6]Precision = TP/(TP+FP); [7]Recall = TP/(TP+FN); [8]F1 score = 2TP/(2TP+FP+FN); [9]Absolute difference between true and inferred, averaged across time; [10]Absolute difference between true and inferred fitness averaged over true positive lineages. *indicates the best performance.

as adapted, our fitness estimates $\hat{s}_i$ are both accurate and precise in both regimes. The full comparison between methods is provided in Table 1).

While the performance characteristics of our method are comparable or superior to those of the existing approaches, BASIL nevertheless fails to identify a significant proportion (between 15 and 20% in our simulations) of truly adapted lineages (Table 1). This is probably a biological inevitability rather than a flaw of the method itself. Some lineages, particularly those with weak adaptive mutations, may simply not reach large enough population sizes before they begin to decline due to clonal interference, which would prevent them from being identified as adapted by any method. Indeed, we find that false negative lineages in our simulations are typically only slightly more fit than the ancestor and their trajectories are essentially indistinguishable from those of truly neutral lineages, which supports our conjecture (see Section 4.2 in the Supplementary Information and Figures S6–S8 for more details).

Comparison of methods on published BLT data

We next sought to compare the performance of BASIL against FitMut2 on data from published BLT experiments. In the absence of ground truth, we took two approaches to carry out this comparison. First, Venkataram et al sampled 410 yeast clones from their BLT experiments and estimated their fitness in two separate competition assays, carried out in the presence and absence of the alga *Chlamydomonas reinhardtii* [33]. Although the fitness of these clones may differ from the fitness of the corresponding adapted lineages in the original BLT experiments, these estimates represent the best available approximation of the ground truth. Thus, we asked how accurately BASIL and FitMut2 predict the fitness of these clones based on the frequency trajectories of the corresponding lineages in the BLT experiments. Second, we evaluated how the performance of FitMut2 and BASIL degrades as we subsample the Levy et al data [3], which is the largest and the "cleanest" BLT dataset available so far (see Data S1, Tab 3 and Figure 2E).

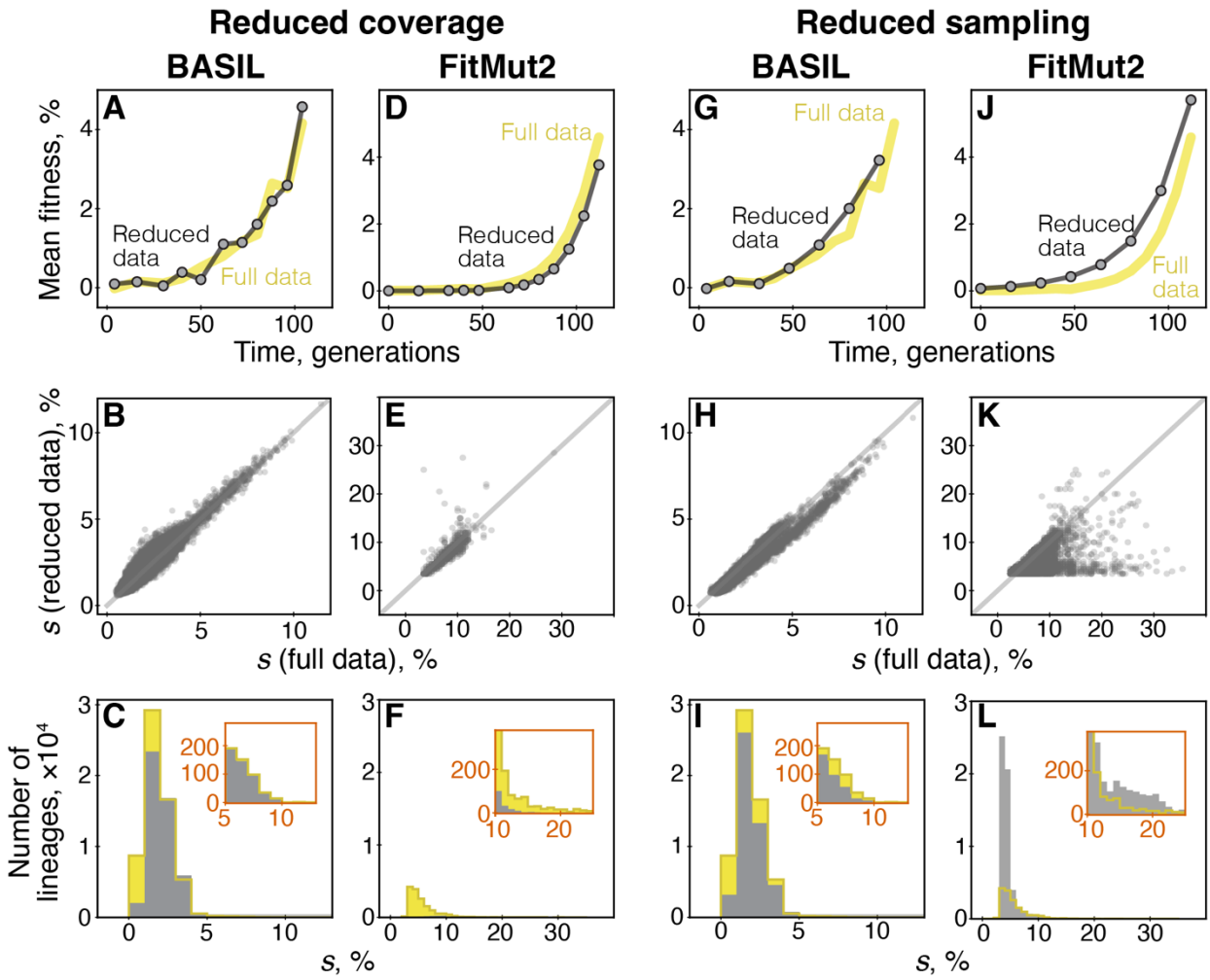

Figure 6. The effects of down-sampling on BASIL and FitMut2 performance. We reanalyzed the Levy 2015 dataset (Replicate 1) using BASIL and FitMut2 after reducing sequencing coverage or sampling frequency. See Materials and Methods for details. **A.** Mean fitness trajectories inferred by BASIL on full and reduced data. **B.** Inferred selection coefficients for lineages identified as adapted by BASIL in both full and reduced datasets. The 1-to-1 line is shown in light gray. **C.** Distribution of fitness effects of lineages identified as adapted by BASIL in full and reduced data. Inset shows the tail of the distribution. **D–F.** Same as A–C but for FitMut2. **G–L.** Same as A–F but for reduced sampling.

Comparison with competition assay measurements in Venkataram 2023 data. We find that the Pearson correlation coefficient between the fitness of yeast clones measured in competition assays and the corresponding BLT fitness estimates produced by BASIL is 0.751 and 0.636 (Figure S9) for the experiments with and without the alga, respectively; and the average difference between the estimates is 2.56% and 3.52%. For FitMut2, the Pearson correlation coefficients in the two environments are 0.257 and 0.47 and the average difference between the estimates is almost three times larger than for BASIL, 6.02% and 10.34%, respectively.

Down-sampling analysis on Levy 2015 data. To test how the amount of data affects BASIL and FitMut2 performance, we down-sampled the Levy 2015 dataset [3] in two ways (see Materials and Methods). First, we reduced average coverage from about 160× to about 10× per lineage. Second,

we reduced the sampling frequency from 12 time points to 8. We then compared how various performance metrics for each method degrade relative to the full dataset.

The results of this analysis are shown in Figure 6. We find that when coverage is reduced, the mean fitness trajectories inferred by either method remain virtually unchanged (Figure 6A,D) but the number of lineages identified as adapted decreases. Specifically, BASIL identifies 49,330 lineages as adapted in the reduced data compared to 60,920 in the full data, a 19% decrease. In contrast, FitMut2 identifies only 1,131 lineages as adapted in the reduced data compared to 15,386 in the full data, a 92.6% decrease. For both methods, the overwhelming majority (≥ 98.8%) of lineages identified as adapted in the reduced data are also identified as such in the full data, and the estimates of selection coefficients remain largely unchanged (Pearson correlation coefficient of 95.8% and 82% for BASIL and FitMut2, respectively; and an average change of 0.20% and 0.83% in the estimated value of s, respectively; Figure 6B,E and Table S7). Notably, for BASIL, lineages that are no longer identified as adapted after down-sampling are those with small selection coefficients whereas FitMut2 loses adapted lineages even with large selection coefficients (Figures 6C,F and Table S7).

When we reduce sampling frequency, BASIL exhibits a similarly modest degradation in performance. The number of adapted lineages decreases by 21% (from 60,920 to 48,206), with 99.9% calls being the same as in the full dataset. Their inferred fitness effects also remain consistent (Pearson correlation coefficient of 97.9% and 0.18% change in s, Figure 6H and Table S7). One noteworthy difference compared to reduction of coverage is that, when sampling is reduced, BASIL loses power to identify lineages with both weak and strong adaptive mutations (compare panels C and I in Figures 6).

In contrast to BASIL, FitMut2 exhibits a dramatic degradation of performance under sampling frequency reduction. Pathologically, in addition to 15,386 lineages identified as adapted in the full data, FitMut2 finds 39,862 additional likely spurious adapted lineages in the reduced data, most of which have small selection coefficients (Figure 6L and Table S7). Moreover, for those lineages that are identified as adapted in both full and reduced data, FitMut2 estimates different selection coefficients (Pearson correlation coefficient of 62.0% and 1.29% change in s, Figure 6K and Table S7).

In summary, reduction in sampling frequency is more detrimental for inference than a uniform reduction in coverage across all time points for both methods. Yet, while both types of data reduction degrade BASIL's performance only to moderate degree, FitMut2 exhibits a dramatic loss of power under coverage reduction and a troubling increase in the rate of false positives under sampling frequency reduction. Thus, taken together, these analyses demonstrate that BASIL is a more statistically robust, accurate and precise approach for the analysis of BLT data than FitMut2.

Application of BASIL to data from BLT experiments

Finally, we applied BASIL to five published BLT experiments (including Levy et al [3] and Venkataram et al [33]) carried out with various strains of yeast *S. cerevisiae* propagated in different environmental conditions (see Data S1, Tab 3). The main goal of this analysis was to expose BASIL to a variety of evolutionary scenarios, examine its outputs and look for potential failure modes that we have not observed so far. To this end, we analyzed BLT experiments described in Refs. [3,4,29,32,33].

We find that the mean fitness inferred by BASIL is generally consistent across replicates and with a posteriori estimates calculated based on the inferred selection coefficients and frequencies of

identified adapted lineages (Figure S10, Data S1, Tab 3; see Methods). However, in the Venkataram 2023 data, we find that the mean fitness inferred by BASIL begins to decline at later time points. Since there is a strong theoretical expectation that the mean fitness in large populations should not decline, this observation indicates that BASIL fails to accurately infer mean fitness at later time points in this dataset. We suspect that this failure is caused by the prevalence of secondary adaptive mutations, i.e., adaptive mutations arising in already adapted lineages [4,5]. Given that adaptation in this experiment appears to be extremely fast (0.28% per generation on average;

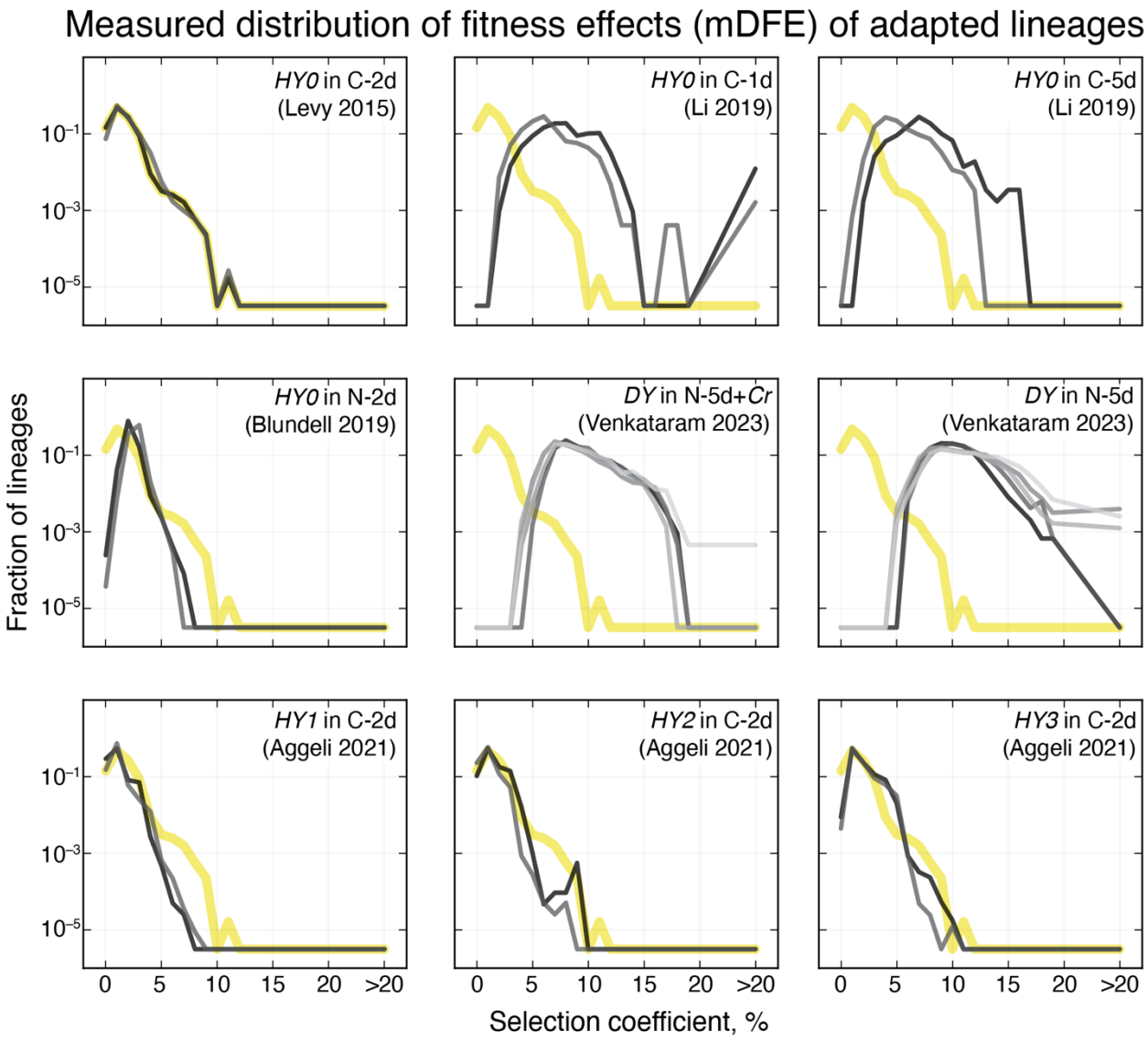

Figure 7. Histograms of the measured distributions of fitness effects (mDFEs) of adapted lineages inferred by BASIL in nine BLT experiments. Each panel shows a BLT experiment, as indicated. Gray lines are different replicates. In the top right corner, we provide a short-hand description of the strain (haploids are denoted HYi, diploids are denoted by DY) and environment used in each experiment. "C" and "N" refer to the likely limiting nutrient in the medium, -xd indicates the number days in the growth and dilution cycle, and "+*Cr*" indicates the presence of the alga *Chlamydomonas reinhardtii* in the evolution environment (see Data S1, Tab 3 for additional details). Thick yellow line shows the mDFE of *HY0* in C-2d (Levy 2015, Replicate 1) as reference.

Data S1, Tab 3), secondary adaptive mutations likely arise sooner than in other experiments. How secondary mutations can lead to declines in the inferred mean fitness can be rationalized as follows. As lineages acquire secondary adaptive mutations, their frequencies begin to rise faster than expected from previous observations. However, at later time points, BASIL has high confidence in the fitness estimates of most lineages. Thus, as it estimates the mean fitness at the next time interval, it attributes the unexpected frequency increases of multiple lineages to a decline in the mean fitness. We note that, while BASIL's apparent inability to accommodate changes in belief distributions at later stages of the BLT time course is sub-optimal, it does not represent a major problem. Indeed, the main purpose of BLT experiments is to measure the fitness effects single-mutant neighbors of a given wildtype genotype and to isolate some of them for further investigation [3,29,32,33,37,38]. Thus, the focus on typical BLT experiments is on the early stages of adaptation before secondary mutations become prevalent (but see Refs. [4,5] that examine longer-term dynamics).

Next, we examine the fitness effects of individual adaptive mutations. We find that the number of identified adapted lineages can vary across replicate BLT experiments by up to threefold (Data S1, Tab 3). We also find that, while the measured distributions of fitness effects (mDFEs) look similar across replicate experiments (Figures 7, S11), they are statistically distinguishable by several measures (Table S6). These differences could arise from real biological variation, measurement noise, or both. As discussed above, variation in coverage and sampling frequency leads to variation in statistical power to detect beneficial mutations (see Section "Comparison of methods on published BLT data"). Our analysis showed that this variation in power is modest in the Levy 2015 data and would not be sufficient to fully explain the differences between the numbers of detected adaptive mutations across replicates. Biologically, strong rare beneficial mutations could arise in different replicates at different times, causing real biological variation in the mean fitness trajectories (Figure S10), which, even if small, could in turn lead to variation in the probability of establishment of weaker beneficial mutations. These observations highlight the difficulties of inferring the shape of the true underlying DFE from the fitness identified adapted lineages.

DISCUSSION

In this work, we showed that the existing approaches to the analysis of BLT data have significant shortcomings. Using simulated data with pre-existing mutations, we found that both the original Levy-Blundell (LB) method and its improved version implemented in the FitMut2 package, can produce biased estimates of the fitness effects of new mutations, particularly under strong selection. We provided the likely mechanistic explanation for the suboptimal performance of the LB method. Using data from the Levy et al BLT experiment [3], we showed that FitMut2 can both exhibit significant loss of statistical power and elevated rates of false positives when the amounts of available data are reduced.

To overcome these shortcomings of existing approaches, we developed BASIL, a Bayesian filtering method for identifying barcoded lineages that carry nascent adaptive mutations and for inferring their selection coefficients. BASIL estimates belief distributions for these selection coefficients and the corresponding lineage sizes, updating them over time based on the observed barcode read counts. Measurement noise is a key feature of our model, and our new experimental data suggests that the measurement variance in the read count scales quadratically with the expected read count, as opposed to a linear scaling that has been postulated in previous studies. This scaling (with ϵ being a free parameter) is implemented in BASIL, and we suggest to use this scaling in future barcode-analysis models. One notable advantage of our framework is that it estimates mean

fitness based on reference lineages, without the need to be designated neutral reference lineages a priori. This makes our approach applicable to a wide range of evolutionary scenarios, including those with strong selection.

When we applied BASIL to real BLT data, we found that, in contrast to FitMut2, BASIL exhibits only a moderate loss of power when the sequencing coverage or sampling frequency are reduced, and its estimates of selection coefficients of adaptive mutations remain consistent.

BASIL computes a time-dependent joint belief distribution for the selection coefficient and size of every lineage. So far, we did not utilize all the information contained in this distribution, but instead chose to classify lineages into adapted and neutral with a simple linear classifier that is based on a snap-shot of this distribution. In our simulations, this simple classifier with the confidence factor of 3.3 (which controls the slope of the classification line) was very effective. How effective it is on real data, which is arguably more complex than our simulations, is difficult to ascertain. We visualize our classification on two-dimensional diagnostic plots where the mean of the fitness belief distribution is plotted against its standard deviation (Figures S5, S12). In some datasets, these plots reveal clustering of lineages, with some clusters clearly containing highly adapted lineages (Figure S12), similar to our simulations (Figure S5). It is possible to isolate these highly adapted clusters by adjusting the confidence factor, which can be easily done by the user in the current implementation of BASIL. However, while such an adjustment would likely improve precision, it may significantly worsen recall, since not all adapted lineages belong to the highly adapted clusters, as we saw in our simulations (Figure S5). Instead, developing more sophisticated classifiers and testing them on more realistic simulations could be a valuable direction for future research.

Our approach has certain limitations. On the technical side, BASIL is quite computationally expensive since it requires extensive Markov Chain Monte Carlo (MCMC) sampling to estimate posterior distributions of lineage fitness and sizes. For instance, the analysis of a BLT experiment with 5×10^5 lineages sampled at 10 time points took us ~50 hours of running time on AMD Ryzen 5 7600X 6-Core Processor, and the running time increases linearly in the number of lineages and sampling time points. However, since MCMC computations for different lineages are independent, the algorithm is easily parallelizable. In our analyses, we found it convenient to use 12–32 processor cores.

One important conceptual limitation of our model is that it assigns a single constant selection coefficient to each barcoded lineage, but many barcoded sub-populations are in fact polymorphic. Indeed, each new adaptive mutation arises in a single cell, and the mutant lineage initially constitutes only a small fraction of the parental barcoded sub-population [3]. The mutant lineage outcompetes the wildtype within this sub-population gradually, such that the barcode frequency increases slower than would be expected from the mutant's selection coefficient alone. Since BASIL ignores these intra-lineages dynamics, it probably underestimates the selection coefficient of lineages that arise later in the BLT experiment and also undercounts them. However, the magnitude of this bias is hard to predict a priori because it depends on the rate of adaptive mutations and their fitness effects, the very quantities that we are trying to estimate.

A barcoded sub-population also becomes polymorphic when secondary adaptive mutations arise and begin to establish [5]. In principle, BASIL is capable of updating the belief distribution for any given lineage if its fitness changes, but in practice, convergence to a stable new distribution is probably slow at later time points when the confidence in the current belief is already high. As a

result, when secondary mutations become dominant, BASIL can fail. While this represents a limitation, it is not a major one since the purpose of typical BLT experiments is to estimate the selection coefficients of single adaptive mutations rather than their combinations. Thus, when secondary beneficial mutations become common, BLT experiments become much less informative and are typically aborted. But given this limitation, BASIL is currently not expected to be reliable in BLT experiments under fluctuating or frequency-dependent selection or in spatially structured populations where expansion of adapted lineages is sub-exponential [55]. However, it should be possible to extend our approach to these more general cases.

It is important to keep in mind that BASIL only identifies adapted lineages and infers their selection coefficients but does not infer the rate at which beneficial mutations arise. To infer the latter, we would need to know not only how many beneficial mutations were detected but also how many arose but were lost by drift while rare and how many were established but not detected. In other words, we would need a model of mutant's evolutionary dynamics as well as a model of sampling and detection. Previously, Levy et al developed the former model and inferred the rates of adaptive mutations with different fitness effects [3], but their model did not account for biases introduced by sampling and detection. Our results suggest that these biases may not be negligible (Figures 6 and 7 and Table S6). Therefore, incorporating them into a model of beneficial mutation rate inference is an important problem for future research.

MATERIALS AND METHODS

Code and data availability

Code. All analysis code is written in Python and is available at https://github.com/Huanyu-Kuo/BASIL-public. Computationally intensive analyses were run on the Triton Supercomputing Cluster (TSCC).

Data. Barcode sequencing data generated in this work are available at https://data-view.ncbi.nlm.nih.gov/object/PRJNA1227420.

Estimation of noise in the barcode frequency measurement process

Strains. 26 clones were isolated from the barcoded library of the diploid strain GSY6699 of yeast *Saccharomyces cerevisiae* used in our previous work [33]. Details of the strain and library construction are provided in Refs. [3,31]. Briefly, each clone carries a unique 30-bp DNA barcode that replaces one copy of the YBR209W locus; all clones are otherwise identical in the rest of the genome. Barcode sequences of all clones have been determined previously [33] and are reported in Data S1, Tab2.

Culture preparation. To probe the measurement noise process, we aimed to construct a barcoded clone mixture consisting of 5 clones at each of five frequencies 0.1, 0.01, 10^{-3}, 10^{-4}, 10^{-5}, and one clone at the remaining frequency of approximately 0.45 (26 clones total). We started a liquid culture of each clone from frozen stocks in 3 ml of YPD (10 g yeast extract, 20 g peptone, 20 g dextrose, 1 L Milli-Q water) in 16-mm test tubes, and incubated these cultures overnight, in a shaking incubator set at 200 rpm at 30°C. After overnight growth, culture densities were measured with a Coulter Counter and were found to be around 1.5×10^{8} cell/ml. The measured values for all cultures are provided in Data S1, Tab2. For clones with the four lowest target frequencies (those below and including 0.01), we diluted the overnight cultures 1:10, 1:100, 1:10^3 and 1:10^4 in PBS, respectively. To make the final clone mixture, we pooled 200 μl of each diluted culture, then added

200 µl of the five cultures with target frequency 0.1, and then added 600 µl of the culture with the largest frequency. This resulted in final mixture volume of 5.6 ml at density approximately 4.6×10^7 cell/ml with barcodes clustered around six distinct frequencies: 0.395, 0.108, 0.019, 1.24×10^{-3}, 1.28×10^{-4} and 1.41×10^{-5}. Our estimates of clone frequencies in the mixture are provided in Data S1, Tab2. We spun down the clone mixture, re-suspended it in 20 ml of PBS (final density 1.57×10^7 cell/ml, split it into 20 aliquots (1 ml each in a 2 ml cryogenic tube), supplemented with 500 µl of 80% glycerol, and stored at –70°C.

Barcode sequencing. We followed the same protocols for DNA extraction and sequencing library preparation as in our previous work [33]. We repeated this procedure 10 times, each time using one aliquot of the barcoded clone mixture stocks described above. All ten replicate libraries were sequenced on the Illumina HiSeq platform, resulting in approximately 2.3×10^{-6} total reads.

Data analysis. We used BarcodeCounter2 to identify barcodes in the sequencing data and count them [56]. The data are shown Data S1, Tab2. One replicate was excluded from analysis due to low coverage. In addition, we could not identify barcodes for two out of 26 clones, and two other clones (in the lowest frequency class) were not detected in the sequencing data. These four barcodes were removed from further analysis. Thus, we retained 22 barcodes present in all $n_{\text{rep}} = 9$ replicates.

Denote by X_p the set of barcodes b with the same target frequency p and denote the read count for barcode b in replicate i by r_{bi}. For each frequency class X_p and replicate i, we estimate the variance in the read count due to measurement noise as $\hat{\sigma}_{pi}^2 = \frac{1}{|X_p|}\sum_{b\in X_p}(r_{bi} - \bar{r}_{bi})^2$, where $\bar{r}_{bi} = \frac{R_i}{n_{\text{rep}}}\sum_{j=1}^{n_{\text{rep}}}\frac{r_{bj}}{R_j}$ is the estimate of the expected read count for barcode b in replicate i, R_i is the total coverage in replicate i and $|X_p|$ denotes the total number of barcodes in the frequency class p. More details are provided in Section 1 in the Supplementary Information.

We model the relationship between read count mean $\bar{r}_{bi}$ and variance $\hat{\sigma}_{pi}$ shown in Figure 4B as $\hat{\sigma}_{pi}^2 = a\bar{r}_{pi} + \epsilon\bar{r}_{pi}^2$ and test four special cases of this equation. To recapitulate the Poisson distribution, we set $a = 1$ and $\epsilon = 0$. We refer to the noise model proposed by Levy et al (see equation S8) as "Linear general". In this model, $\epsilon = 0$ and a is a free parameter. To capture the possibility that the degree of overdispersion varies across read depths, we test a quadratic relationship with $a = 1$ and a fitting parameter $\epsilon > 0$, which matches the Poisson model for small $\bar{r}_{bi}$. This is the model in equation (2), and we refer to it as "Quadratic with 1-parameter". Finally, we also fit the general quadratic model with two fitting parameters a and ϵ.

We find the best-fit parameters using the non-linear least squares method. Since the estimated measurement noise variances span six orders of magnitude (see Figure 4B), we carry out the fitting in the log space, i.e., we minimize the squared error between $\log\log\hat{\sigma}_{pi}^2$ and $\log a\bar{r}_{pi} + \epsilon\bar{r}_{pi}^2$. In Table S1 we report the best-fit parameters, residual sums of squares (RSS), as well as the coefficient of determination R^2, computed as 1–RSS/TSS, where TSS denotes the total sum of squares. We compare the goodness-of-fit for all pairs of nested models using the F-test. Specifically, assuming that the null model has p_{null} free parameters and the alternative model has p_{alt} parameters $p_{\text{null}} < p_{\text{alt}}$, we calculate the F-statistic as $F = \frac{\text{RSS}_{\text{null}} - \text{RSS}_{\text{alt}}}{\text{RSS}_{\text{alt}}}\frac{n - p_{\text{alt}}}{p_{\text{alt}} - p_{\text{null}}}$, where n denotes the number of data points, and RSS_{null} and RSS_{alt} are the residual sums of squares for the null and alternative model, respectively. If the data is generated by the null model, then the F-statistic is distributed according to the F-distribution with $(p_{\text{alt}} - p_{\text{null}}, n - p_{\text{alt}})$ degrees of freedom.

The results of this analysis are shown in Figure 4B and Tables S1 and S2. We find that the quadratic models explain at least 93% of variance in our data compared to linear models, which explain 89% or less. We also find that the general quadratic model provides only a marginally better fit than the 1-parameter quadratic model, with a *P*-value of only 0.0235. Therefore, we use the 1-parameter quadratic model as our model for the relationship between read count mean and variance.

BLT simulations

We carry our simulations in the weak and strong selection regimes (Figure S1). We chose the parameters for the weak selection regime to make it similar to the experimental conditions described by Levy et al [3]. In particular, selection and genetic drift are relatively weak, and the variance in the read count is a linear function of the mean. For the strong selection regime, we chose parameters that should make inference harder. In particular, selection and drift are stronger, and the variance in the read count is a quadratic function of the mean (see equation (2)). Simulation parameters for both regimes are provided in Table S3.

All simulations start with the same number $N_L = 10^5$ of barcoded lineages. The initial size of each lineage is drawn from an exponential distribution with mean $\langle N \rangle / N_L$, where $\langle N \rangle$ is the expected total population size (see Table S3). In each simulation, 3,000 randomly chosen lineages are adapted, such that all members of such lineages have the same beneficial mutation. We draw these lineages uniformly with the constraint that their initial size must exceed $1.5D = 384$ individuals, where $D = 256$ is the dilution factor. The selection coefficient (fitness) s_i of each adapted lineage is drawn from the normal distribution with mean $\langle s \rangle$ and standard deviation σ_s, and negative draws are discarded. The fitness of all other lineages is set to $s = 0$. No new mutations arise during the simulation.

After initializing our populations, we simulate their growth and dilution, with each growth-dilution cycle consisting of the following steps:

1. Each cycle begins with the calculation of the population's mean fitness $\bar{s} = \sum_i s_i n_i / N$, where n_i is the current size of lineage i and $N = \sum_i n_i$ is the current population size.
2. We then simulate the dilution process. For each lineage i, we draw its new size n_i' as a random number from the Poisson distribution with mean n_i/D. Thus, after dilution, the population's bottleneck size equals $N_b = \sum_i n_i' \approx \langle N \rangle / D$.
3. We then simulate population expansion. Each lineage i expands by the factor $W_i = D \exp[(s_i - \bar{s})\Delta t]$, where $\Delta t = 8$ generations is the duration of the growth phase. Thus, at the end of the growth phase, the new lineage size is $n_i^{\text{new}} = n_i' W_i$ (rounded to the nearest integer) and the total population size is again $\sum_i n_i^{\text{new}} = D \sum_i n_i'(1 + (s_i - \bar{s})\Delta t) \approx \langle N \rangle$.

Each simulation lasts for a total 20 growth-dilution cycles (160 generations). We simulate sampling and barcode sequencing by selecting R random individuals from the population at the end of every other cycle (i.e., before dilution). For each sampled barcode i, the number of sequencing reads r_i is drawn from the negative binomial distribution with mean $\langle r_i \rangle = n_i R / N$ and variance σ_r^2, where R is the sequencing depth (see Table S3). The output of each simulation is a table of "sequenced" barcode counts at all sampling time points.

The neutral decline method

The neutral decline method retains the essential features of the original LB method but simplifies it in two major ways. First, in the LB model, adapted lineages are characterized by two parameters, the selection coefficient s and the establishment time τ, whereas the neutral decline method has only one parameter s per lineage. However, in our simulations, all adapted mutations are pre-existing, implying that our simpler one-parameter model is sufficient. The second important difference is that, in addition to the mean fitness $\bar{s}_k$, the LB method also estimates the noise parameter κ (see Section 2 in the Supplementary Information). To infer both $\bar{s}_k$ and κ, the LB method uses a model that describes how the entire distribution of read counts changes over time (equations (S8), (S9)). The mean of this distribution is described by equation (1), which does not depend on κ. Thus, equation (1) is sufficient if we are interested only in $\bar{s}_k$ (see Figure S2). Importantly, since equation (1) directly follows from the LB model, a failure of this equation would imply a failure of the full LB model. Therefore, the focus our analysis on the neutral decline rather than the full LB model is justified.

The neutral decline method proceeds as follows.

1. For each consecutive time interval (t_{k-1}, t_k), we choose the neutral reference lineages as those having certain read counts r_{k-1} at the beginning of the interval with either $20 \leq r_{k-1} \leq 40$ (low-abundance reference) or $80 \leq r_{k-1} \leq 100$ (high-abundance reference).
2. For each initial read count r_{k-1}, we calculate the mean fitness $\bar{s}_k$ using equation (1) and over all initial read counts (for reference lineages only). We repeat the mean fitness calculation for all sampling intervals and thereby obtain the full mean-fitness trajectory $\bar{\boldsymbol{s}} = (\bar{s}_1, \bar{s}_2, \dots)$.
3. To infer the fitness effect s_i of lineage i, we maximize the log-likelihood function of the lineage trajectory $\boldsymbol{r}_i = (r_{i1}, r_{i2}, \dots)$, $\hat{s}_i = \operatorname{argmax} \ell(s)$, where $\ell_i(s) = \sum_k \log p\,(r_{ik}; \mu_k(s, \bar{s}_k), \kappa)$ with $p(r; \mu, \kappa)$ given by equation (S8), and $\mu_k(s, \bar{s}_k)$ given by equation (S9) and $\kappa = 3$.
4. To determine whether lineage i is adapted, we apply the likelihood ratio test to compare the neutral model ($s_i = 0$, null hypothesis) and the adaptive model ($s_i = \hat{s}_i$, alternative hypothesis). The likelihood-ratio test statistic is given by $R_i = -2(\ell_i(0) - \ell_i(\hat{s}_i))$. We call the lineage adapted if $\hat{s}_i > 0$ and $R_i \geq 3.84$, which is the 95th percentile of the χ^2-distribution with 1 degree of freedom.

Empirical test of the neutral decline equation

In Figure 2, to test the neutral decline equation (1), we use the barcode read count data from two BLT simulations and the three BLT experiments Levy 2015 (replicate 1 in Ref. [3]), Li 2019 (replicate 2 from the 1-day transfer cycle experiment from Ref. [29]), and Venkataram 2023 (replicate 5 of the experiment with the alga from Ref. [33]). For each real or simulated BLT experiment, we selected one "early" and one "late" sampling time interval (t_{k-1}, t_k), as described in the caption to Figure 2. For each interval, we calculate the average read count $\bar{r}_k$ and the rescaled average read count $\tilde{r}_k$ for lineages that had the same read count $r_{k-1} \in [0,100]$. If the number of lineages with a given r_{k-1} was less than 5, we did not estimate the corresponding $\bar{r}_k$. To investigate whether $\tilde{r}_k$ is a linear function of r_{k-1} as predicted by equation (1) for neutral lineages, we plot $\tilde{r}_k$ against r_{k-1}. For some r_{k-1} in some datasets, this dependence becomes very noisy or clearly deviates from linearity. Thus, we first visually identify the intervals of r_{k-1} where the data behaves approximately linearly. We choose the interval $5 \leq r_{k-1} \leq 15$ for the Li 2019, Venkataram 2023 and the strong-selection simulated datasets at the late time point; we choose the interval $5 \leq r_{k-1} \leq 40$ for the Venkataram 2023 dataset at the early time point; and we choose $5 \leq r_{k-1} \leq 99$ for all the other

datasets. Then, using the data only within these intervals, we fit a linear model $\tilde{r}_k \sim ar_{k-1} + b$ either with $b = 0$ (restricted model) or with b being a free parameter (full model) and ask whether the full model provides a statistically better fit to our data using the F-test (see Section "Estimation of noise in the barcode frequency measurement process").

BASIL

The detailed description of BASIL is given in Section 3 of the Supplementary Information. Here, we provide the key expressions and brief descriptions of the algorithm.

Model of measurement. We model the probability of observing r reads for a lineage with n cells in a total population of size N which was sequenced to depth R as

$$P^{\text{meas}}(r|n;\epsilon) = \frac{\Gamma(r+1/\epsilon)}{\Gamma(r+1)\Gamma(1/\epsilon)}\left(\frac{\epsilon\langle r\rangle}{1+\epsilon\langle r\rangle}\right)^r\left(\frac{1}{1+\epsilon\langle r\rangle}\right)^{\frac{1}{\epsilon}}, \tag{5}$$

where $\langle r\rangle = nR/N$ is the expected number of reads.

The belief distribution for lineage size and fitness. Our main goal is to estimate the belief probability distributions $P_{ik}^{\text{belief}}(n,s) \equiv P^{\text{belief}}(n,s|\boldsymbol{r}_{ik})$, where $\boldsymbol{r}_{ik} = (r_{i0},\dots,r_{ik})$ is the read count vector for lineage i up to and including sampling time point t_k. We express these belief distributions as $P^{\text{belief}}(n,s|\boldsymbol{r}) = P^{\text{belief}}(n|s,\boldsymbol{r})\,P^{\text{belief}}(s|\boldsymbol{r})$ with parametric forms $P^{\text{belief}}(n|s,\boldsymbol{r}) = \gamma\big(n;\varkappa(\boldsymbol{r}),\theta(s,\boldsymbol{r})\big)$ and $P^{\text{belief}}(s|\boldsymbol{r}) = N(s;\mu(\boldsymbol{r}),\sigma^2(\boldsymbol{r}))$, where $N(s;\mu,\sigma^2)$ is the normal distribution with mean μ and variance σ^2 and $\gamma(n;\varkappa,\theta)$ is the gamma distribution with the shape parameter $\varkappa$ and scale parameter θ and we use the family of functions

$$\theta(s,\boldsymbol{r}) = \frac{a(\boldsymbol{r})}{\varkappa(\boldsymbol{r})}\exp\left[\frac{b(\boldsymbol{r})\big(s-\mu(\boldsymbol{r})\big)}{\sigma(\boldsymbol{r})}\right].$$

Thus, our belief distributions come from a family with five parameters μ, σ^2, $\varkappa$, a and b all of which are fit based on the data vector $\boldsymbol{r}$.

At t_0, we set $\mu = 0$, $\sigma^2 = 0.1$ and $b = 0$ for all lineages; and for lineage i, we set $\varkappa(r_{i0}) = r_{i0}/(1+0.01r_{i0})$, $a = r_{i0}N/R_0$ where R_0 is the coverage at the initial time point. We believe that the prior distribution should be fairly broad (i.e., uninformative), but this particular choice is likely unimportant.

Projecting the belief distribution. To obtain the prior distribution $P_{ik}^{\text{prior}}(n,s)$ for the sampling time t_k, we project the belief distribution $P_{ik-1}^{\text{belief}}(n,s)$ as follows. First, we obtain the 1-cycle projected probability distribution

$$P_1(n,s) = \int_0^\infty P^{\text{c}}(n|n';s,\bar{s})\,P_{ik-1}^{\text{belief}}(n',s)\,dn',$$

where $P^{\text{c}}(n|n';s,\bar{s})$ is the one-cycle transition probability given by equation (S34). The marginal distribution for s remains unchanged during projection. The probability that the lineage goes extinct during one cycle is $\left(\frac{D}{D+\theta}\right)^{\varkappa}$ where D is the dilution factor and $\varkappa$ and θ are the parameters of the belief distribution $P_{ik-1}^{\text{belief}}(n,s)$. Conditional on lineage survival, we approximate the distribution $P_1(n|s,\boldsymbol{r})$ by the gamma distribution with shape parameter $\varkappa_1$ and scale parameter θ_1, which can be calculated as functions of s, $\varkappa$ and θ (equations (S42) and (S43)). If multiple cycles elapse between successive sampling time points, the prior probability $P_{ik}^{\text{prior}}(n,s)$ is obtained by applying this procedure recursively.

Updating the belief distribution. The belief distribution is updated using Bayes' theorem (equation (3)). We would like to use the same analytical parametric as above for the updated belief

distribution, but since the numerator of this equation is a complex function of both n and s, the normalization constant Z_{ik}^{sel} and the moments of the right-hand side of equation (3) cannot be computed analytically. Therefore, to estimate the parameters of the updated belief distribution, we use the Markov Chain Monte Carlo (MCMC) approach described in Section 3.2.2 in the Supplementary Information.

Calculation of Z_{ik}^{neut}. Equation (4) for the estimation of mean fitness $\bar{s}_k$ and the noise parameter ϵ_k in the time interval (t_{k-1}, t_k) depends on the marginal likelihood Z_{ik}^{neut} that lineage i which was classified as putatively neutral in this interval has r_{ik} reads at time t_k. We obtain Z_{ik}^{neut} as follows. First, we approximate the marginal belief distribution $P_{ik-1}^{\text{belief}}(n)$ for lineage size n at t_{k-1} by a gamma distribution with shape parameter $\tilde{\varkappa} = \varkappa\left[(1+\varkappa)e^{b^2} - \varkappa\right]^{-1}$ and scale parameter $\tilde{\theta} = ae^{\frac{b^2}{2}}/\tilde{\varkappa}$, where $\varkappa$, a and b are the parameters of the original distribution $P_{ik-1}^{\text{belief}}(n,s)$. We then obtain the prior distribution $P_{ik}^{\text{prior}}(n)$ under the neutral model analogously to how we obtain $P_{ik}^{\text{prior}}(n,s)$ in the model with selection. Finally, we use MCMC to estimate Z_{ik}^{neut} (see Section 3.3 in the Supplementary Information for details).

Performance of BLT analysis methods on experimental data.

Down-sampling. To test how sequencing coverage and sampling effort affect our mDFE inference, we down-sampled the data from Replicate 1 of the original BLT experiment with strain HY0 in the C-2d environment [3]. The original data has the average coverage of ~160× per lineage, and sampling occurred at 12 time points, at generations 0, 16, 32, 40, 48, 64, 72, 80, 88, 96, 104 and 112. To reduce coverage, for each barcode whose read count is r_{ik} in the original data, we draw a random number r'_{ik} from a Poisson distribution with mean $r_{ik}/16$, such that the average coverage per lineage is ~10. To reduce sampling frequency, we remove sampling time points at generations 40, 72, 88 and 104, such that the total number of samples is reduced from 12 to 8.

Mean fitness based on inferred adapted lineages. Since the true mean fitness in the BLT experiments is unknown, we follow Levy et al [3] and validate the mean-fitness estimates obtained by BASIL by comparing them to a posteriori mean-fitness estimates based on the inferred selection coefficients and frequencies of identified adapted lineages $\bar{s}_k^{\text{ADP}} = \sum_i \hat{s}_i\, r_{ik}/R_k$ where $\hat{s}_i$ is the inferred selection coefficient of adapted lineage i.

Comparison of replicate mDFEs. BASIL estimates of the selection coefficients of adapted lineages for all datasets shown in Figure 7 are provided in Data S1, Tabs 4–12. To construct the measured distribution of fitness effects of new mutations (mDFE) shown in Figure 7, we group lineages based on their inferred selection coefficient $\hat{s}_i$ into 1-percent bins as well as a bin of all lineages with $\hat{s}_i > 0.2$. The statistics of these mDFEs are reported in Data S1, Tab 3.

We test the consistency of mDFEs across replicates of the same BLT experiment using quantile-quantile (Q-Q) plots (Figure S11). We find that, by this metric, most mDFEs inferred from replicate experiments are consistent with each other, although experiments HY0-C1d [29], HY2-C2d, HY3-C2d [32] and DY-N5d [33] show some discrepancies. In addition, we compare the mean, the median, the interquartile range (IQR), and the interdecile range (IDR) across replicate mDFEs using a permutation test. To this end, we create a list of selection coefficients of all lineages identified as adapted in any of the replicates and then randomly reassign each selection coefficient to each replicate i while preserving the original number of selection coefficients in each replicate. We estimate each mDFE statistic of interest for each permuted replicate. Then, for each pair of

replicates it in the permuted data, we calculate the absolute value of the difference between the statistic values; and we average them over all pairs to obtain our test statistic. We repeat this permutation procedure 5,000 times to obtain the null distribution for each mDFE statistic and calculate the empirical *P*-value based on this distribution. The results of this more sensitive test are reported in Table S6.

ACKNOWLEDGEMENTS

We thank members of the Kryazhimskiy lab for discussions and feedback and anonymous reviewers for constructive feedback on the initial version of the manuscript. The UC San Diego Triton Computing Cluster assisted with computational work. This work was supported by the Career Award at Scientific Interface (1010719.01) from the Burroughs Wellcome Fund to SK, the Alfred P. Sloan Research Fellowship (FG-2017-9227) to SK, the Hellman Fellowship to SK, and the NIH grants 1R01GM137112 and R35GM153242 to SK.

AUTHOR CONTRIBUTIONS

Conceptualization: HK, SK; Experimental design: HK, SK; Experiments: HK; Analysis: HK, SK; Writing: HK, SK.

COMPETING INTERESTS

None

SUPPLEMENTARY MATERIALS

Supplementary Text

Supplementary Figures

Figure S1. Simulated BLT data.

Figure S2. Consistency of mean-fitness inference for BLT data by Levy et al.

Figure S3. Parametric fit to the MCMC-sampled belief distributions.

Figure S4. Dynamics of the belief distribution.

Figure S5. Discrimination of adapted and neutral lineages in simulated data.

Figure S6. The frequency dynamics of lineages in the weak-selection regime, stratified by class label.

Figure S7. The frequency dynamics of lineages in the strong-selection regime, stratified by class label.

Figure S8. Selection coefficients and sizes of lineages identified as adapted in simulated data.

Figure S9. Comparison of BASIL and FitMut2 performance on Venkataram et al data.

Figure S10. Mean fitness trajectories inferred by BASIL in real BLT datasets compared to mean fitness trajectories calculated from adapted lineages.

Figure S11. Quantile-quantile plots for the distributions of selection coefficients of lineages identified as adapted in replicate BLT experiments.

Figure S12. Discrimination of adapted and neutral lineages in real BLT data.

Supplementary Tables

Table S1. Fitted parameters for models of the relationship between read-count mean and variance.

Table S2. Statistical comparison of models of the relationship between read-count mean and variance.

Table S3. Simulation parameters.

Table S4. Summary of BLT analysis methods.

Table S5. Mathematical notations used in BASIL.

Table S6. Differences of mDFE statistics across replicates in real BLT datasets.

Table S7. Effects of down-sampling on the BASIL and FitMut2 inference.

Supplementary Data

Data S1 [Excel]. Aggregated Data Tables.

- Tab 1. Contents.
- Tab 2. Measurement noise experiment.
- Tab 3. Characteristics of published BLT studies and some summary statistics of the BASIL analysis.
- Tabs 4–12. BASIL results for published BLT experiments.

Raw Data Repository

https://dataview.ncbi.nlm.nih.gov/object/PRJNA1227420

Code

https://github.com/HuanyuKuo/BASIL-public

REFERENCES

1. Gerrish PJ, Lenski RE. The fate of competing beneficial mutations in an asexual population. Genetica. 1998;102–103: 127–144.
2. Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, Botstein D, et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. Nature. 2013;500: 571–574.
3. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature. 2015;519: 181–186.
4. Blundell JR, Schwartz K, Francois D, Fisher DS, Sherlock G, Levy SF. The dynamics of adaptive genetic diversity during the early stages of clonal evolution. Nature Ecology & Evolution. 2019;3: 293–301.
5. Nguyen Ba AN, Cvijović I, Rojas Echenique JI, Lawrence KR, Rego-Costa A, Liu X, et al. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. Nature. 2019;575: 494–499.
6. Traverse CC, Mayo-Smith LM, Poltak SR, Cooper VS. Tangled bank of experimentally evolved Burkholderia biofilms reflects selection during chronic infections. Proc Natl Acad Sci U S A. 2013;110: E250-9.
7. Barroso-Batista J, Sousa A, Lourenço M, Bergman M-L, Sobral D, Demengeot J, et al. The first steps of adaptation of Escherichia coli to the gut are dominated by soft sweeps. PLoS Genet. 2014;10: e1004182.
8. Bedhomme S, Perez Pantoja D, Bravo IG. Plasmid and clonal interference during post horizontal gene transfer evolution. Mol Ecol. 2017;26: 1832–1847.
9. Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature. 2014;514: 54–58.
10. Rogers ZN, McFarland CD, Winters IP, Naranjo S, Chuang C-H, Petrov D, et al. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. Nat Methods. 2017;14: 737–742.
11. Illingworth CJR, Mustonen V. Components of selection in the evolution of the influenza virus: linkage effects beat inherent selection. PLoS Pathog. 2012;8: e1003091.
12. Strelkowa N, Lässig M. Clonal interference in the evolution of influenza. Genetics. 2012;192: 671–682.
13. Williams K-A, Pennings P. Drug Resistance Evolution in HIV in the Late 1990s: Hard Sweeps, Soft Sweeps, Clonal Interference and the Accumulation of Drug Resistance Mutations. G3 . 2020;10: 1213–1223.
14. Schiffels S, Szöllosi GJ, Mustonen V, Lässig M. Emergent neutrality in adaptive asexual evolution. Genetics. 2011;189: 1361–1375.
15. Hallatschek O. The noisy edge of traveling waves. Proc Natl Acad Sci U S A. 2011;108: 1783–1787.
16. Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proceedings of the National Academy of Sciences. 2012;109: 4950–4955.

17. Johnson MS, Reddy G, Desai MM. Epistasis and evolution: recent advances and an outlook for prediction. BMC Biol. 2023;21: 120.
18. Lynch M. Evolution of the mutation rate. Trends Genet. 2010;26: 345–352.
19. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8: 610–618.
20. Couce A, Limdi A, Magnan M, Owen SV, Herren CM, Lenski RE, et al. Changing fitness effects of mutations through long-term bacterial evolution. Science. 2024;383: eadd1417.
21. Sane M, Diwan GD, Bhat BA, Wahl LM, Agashe D. Shifts in mutation spectra enhance access to beneficial mutations. Proc Natl Acad Sci U S A. 2023;120: e2207355120.
22. Ardell SM, Martsul A, Johnson MS, Kryazhimskiy S. Environment-independent distribution of mutational effects emerges from microscopic epistasis. Science. 2024;386: 87–92.
23. Jensen JD, Wong A, Aquadro CF. Approaches for identifying targets of positive selection. Trends Genet. 2007;23: 568–577.
24. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev Genet. 2013;47: 97–120.
25. McDonald MJ. Microbial Experimental Evolution – a proving ground for evolutionary theory and a tool for discovery. EMBO Rep. 2019;20: e46992.
26. Martínez AA, Lang GI. Identifying targets of selection in laboratory evolution experiments. J Mol Evol. 2023;91: 345–355.
27. Ralph P, Coop G. Parallel adaptation: one or many waves of advance of an advantageous allele? Genetics. 2010;186: 647–668.
28. Kuo YP, Hu J, Carja O. Clonal interference, genetic variation and the speed of evolution in structured populations. bioRxivorg. 2025. p. 2025.02.22.639636. doi:10.1101/2025.02.22.639636
29. Li Y, Petrov DA, Sherlock G. Single nucleotide mapping of trait space reveals Pareto fronts that constrain adaptation. Nat Ecol Evol. 2019;3: 1539–1551.
30. Jasinska W, Manhart M, Lerner J, Gauthier L, Serohijos AWR, Bershtein S. Chromosomal barcoding of E. coli populations reveals lineage diversity dynamics at high resolution. Nat Ecol Evol. 2020;4: 437–452.
31. Boyer S, Hérissant L, Sherlock G. Adaptation is influenced by the complexity of environmental change during evolution in a dynamic environment. PLoS Genet. 2021;17: e1009314.
32. Aggeli D, Li Y, Sherlock G. Changes in the distribution of fitness effects and adaptive mutational spectra following a single first step towards adaptation. Nat Commun. 2021;12: 5193.
33. Venkataram S, Kuo H-Y, Hom EFY, Kryazhimskiy S. Mutualism-enhancing mutations dominate early adaptation in a two-species microbial community. Nature Ecology & Evolution. 2023;7: 143–154.
34. Chen V, Johnson MS, Hérissant L, Humphrey PT, Yuan DC, Li Y, et al. Evolution of haploid and diploid populations reveals common, strong, and variable pleiotropic effects in non-home environments. Elife. 2023;12. doi:10.7554/eLife.92899

35. Kinsler G, Li Y, Sherlock G, Petrov DA. A high-resolution two-step evolution experiment in yeast reveals a shift from pleiotropic to modular adaptation. PLoS Biol. 2024;22: e3002848.

36. Abreu CI, Mathur S, Petrov DA. Environmental memory alters the fitness effects of adaptive mutations in fluctuating environments. Nat Ecol Evol. 2024;8: 1760–1775.

37. Blundell JR, Levy SF. Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. Genomics. 2014;104: 417–430.

38. Venkataram S, Dunn B, Li Y, Agarwala A, Chang J, Ebel ER, et al. Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. Cell. 2016;166: 1585-1596.e22.

39. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. Am Nat. 1991;138: 1315–1341.

40. Gallet R, Cooper T, Elena S, Lenormand T. Measuring selection coefficients below 10−3: Method, questions, and prospects. Genetics. 2012;190: 175–186.

41. Johnson MS, Martsul A, Kryazhimskiy S, Desai MM. Higher-fitness yeast genotypes are less robust to deleterious mutations. Science. 2019;366: 490–493.

42. Razo-Mejia M, Mani M, Petrov D. Bayesian inference of relative fitness on high-throughput pooled competition assays. PLoS Comput Biol. 2024;20: e1011937.

43. Li F, Mahadevan A, Sherlock G. An improved algorithm for inferring mutational parameters from bar-seq evolution experiments. BMC Genomics. 2023;24: 246.

44. Li F, Salit ML, Levy SF. Unbiased fitness estimation of pooled barcode or amplicon sequencing studies. Cell Syst. 2018;7: 521-525.e4.

45. Li F, Tarkington J, Sherlock G. Fit-Seq2.0: An improved software for high-throughput fitness measurements using pooled competition assays. J Mol Evol. 2023;91: 334–344.

46. McGee RS, Kinsler G, Petrov D, Tikhonov M. Improving the accuracy of bulk fitness assays by correcting barcode processing biases. Mol Biol Evol. 2024;41: msae152.

47. Weinreich DM. The foundations of population genetics. London, England: MIT Press; 2023.

48. Kuo H. BASIL. https://github.com/huanyukuo/basil-public. 2025. Available: https://github.com/HuanyuKuo/BASIL-public

49. Särkkä S, Svensson L. Bayesian filtering and smoothing. 2nd ed. Cambridge, England: Cambridge University Press (Virtual Publishing); 2023. Available: https://books.google.com/books?hl=en&lr=&id=utXBEAAAQBAJ&oi=fnd&pg=PP1&dq=BAYESIAN+FILTERING+AND+SMOOTHING+Sarkka&ots=GX35LN8qVO&sig=to0oOdHxa10xyWMIqb_ogE__49Y

50. Johnson MS, Venkataram S, Kryazhimskiy S. Best Practices in Designing, Sequencing, and Identifying Random DNA Barcodes. J Mol Evol. 2023. doi:10.1007/s00239-022-10083-z

51. Hashimoto TB, Edwards MD, Gifford DK. Universal count correction for high-throughput sequencing. PLoS Comput Biol. 2014;10: e1003494.

52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

1024 53. Zhang X, Mallick H, Tang Z, Zhang L, Cui X, Benson AK, et al. Negative binomial mixed
1025 models for analyzing microbiome count data. BMC Bioinformatics. 2017;18: 4.

1026 54. Puelma Touzel M, Walczak AM, Mora T. Inferring the immune response from repertoire
1027 sequencing. PLoS Comput Biol. 2020;16: e1007873.

1028 55. Fisher RA. The wave of advance of advantageous genes. Ann Eugen. 1937;7: 355–369.

1029 56. Venkataram S, Kinsler G. BarcodeCounter2. 2021. Available:
1030 https://github.com/sandeepvenkataram/BarcodeCounter2

Supplementary Materials for

Putting BASIL in a BLT: A Bayesian Filtering Method for Estimating the Fitness Effects of Nascent Adaptive Mutations

Huan-Yu Kuo, Sergey Kryazhimskiy*

*E-mail: skryazhi@ucsd.edu

Contents

1 Estimation of measurement noise variance in the calibration experiment **3**

2 Existing methods for the analysis of BLT data **4**
2.1 Review of the Levy-Blundell method 4
2.2 Review of FitMut2 6
2.3 Key assumptions of the Levy-Blundell method 7
2.3.1 Assumption 1. Low-abundance lineages are neutral 8
2.3.2 Assumption 2. Read frequency dynamics faithfully represent lineage frequency dynamics 8

3 BASIL **12**
3.1 Model of measurement 12
3.2 The belief distribution for lineage size and fitness 13
3.2.1 Projecting the belief distribution 13
3.2.2 Updating the belief distribution 16
3.3 Estimating populations mean fitness and the noise parameter 20
3.4 Identification of adapted lineages and estimation of their selection coefficients 23
3.5 Implementation 24

4 Application of BASIL to simulated data **26**
4.1 Belief distribution and lineage identification 26
4.2 Performance on simulated data 26

A Expectation of the noise variance estimator and bias correction **28**

B Expected frequency given the number of reads **29**

C Expectation of the logarithm of a gamma-distributed random variable **31**

5 Supplementary Figures **32**

6 Supplementary Tables **45**

1 Estimation of measurement noise variance in the calibration experiment

Denote by X_p the set of barcodes with the same target frequency p. Variation in the number of reads r_{bi} across barcodes $b \in X_p$ and replicates i arises from three sources. First, the same barcode is represented by different numbers of reads in different replicates since replicates receive different total coverage. Specifically, if the barcode's true frequency in the pool is p_b, we expect to obtain $p_b R_i$ reads for this barcode in replicate i with total coverage R_i. Second, the actual number of reads obtained for barcode b will deviate from $p_b R_i$ by some random amount δr_{bi} due to noise during library preparation and sequencing [2]. We refer to δr_{bi} as "measurement noise". Finally, while we aimed to seed the barcoded clones within the same frequency class X_p with exactly the same target frequency p, in reality, the frequency p_b of barcode in the pool actually deviates from p by some amount δp_b, i.e., $p_b = p + \delta p_b$. We refer to δp_b as the "experimental noise". Putting all of this together, we can model the read count r_{bi} for barcode $b \in X_p$ as

$$r_{bi} = p_b R_i + \delta r_{bi}, \tag{S1}$$

where p_b is unknown and we assume that all δr_{bi} are independent random variables with zero mean, and all δr_{bi} with the same b are identically distributed with variance σ_{pi}^2 (the noise variance) where index p indicates that noise variance can vary between frequency classes X_p, and index i indicates that it can also vary between replicates i. Our goal is to estimate σ_{pi}^2.

The challenge with estimating σ_{pi}^2 is that each barcode has an unknown expected number of reads $p_b R_i$ in each replicate i. If all barcodes had exactly the same frequency p, then we could use r_{bi} from different barcodes in the same replicate to estimate $p_b R_i$ as

$$\bar{r}_{pi} = \frac{1}{|X_p|} \sum_{b \in X_p} r_{bi}. \tag{S2}$$

However, different barcodes $b \in X_p$ in fact have different frequencies p_b, and the estimate (S2), which ignores this variation, would provide a poor estimate for the expected barcode count $p_b R_i$ and risks inflating our estimate of the measurement noise. To overcome this problem, we first estimate p_b as

$$\hat{p}_b = \frac{1}{n_{\text{rep}}} \sum_{j=1}^{n_{\text{rep}}} \frac{r_{bj}}{R_j}.$$

Then, we estimate the expected number of reads $\bar{r}_{bi}$ for barcode b in replicate i as

$$\bar{r}_{bi} = \hat{p}_b \, R_i = \frac{R_i}{n_{\text{rep}}} \sum_{j=1}^{n_{\text{rep}}} \frac{r_{bj}}{R_j}. \tag{S3}$$

We can now use expectation (S3) to estimate the variance in the read count due to measurement noise in frequency class X_p and replicate i as

$$\hat{\sigma}_{pi} = \frac{1}{|X_p|} \sum_{b \in X_p} \left(r_{bi} - \bar{r}_{bi}\right)^2 . \tag{S4}$$

In Appendix A, we derive the expectation for this estimator. We find that this estimator is biased, approximately by a factor $1 - 1/n_{\text{rep}}$ and derive a bias-corrected estimate (see equation (S73)). However, this corrected estimate is not guaranteed to be non-negative, and indeed we found that the variance estimate for one of the replicates in the lowest frequency class becomes negative upon applying this correction. We also found that for 36 out of 41 (88%) values of $\hat{\sigma}_{pi}$ that we could estimate, the correction changes the estimated value by less than 25%. For these reasons, we decided to use the uncorrected estimate given by equation (S4) for the subsequent analysis.

2 Existing methods for the analysis of BLT data

Levy et al developed the first computational method for the analysis of barcode lineage tracking (BLT) data [3], which we refer to as the "Levy-Blundell" method, or the "LB method" for short. More recently, Li et al developed its extension, called FitMut2, and showed that it is superior to the LB method [4]. In this section, we briefly review both the original LB method and FitMut2.

2.1 Review of the Levy-Blundell method

The LB method is based on a stochastic birth-death process that describes a mutant lineage starting from n_0 founding cells that randomly divide and die with certain per capita birth rate b and death rate d. Levy et al derived the full stochastic expression for the probability $P_t(n)$ that the number of cells in the mutant lineage after time t is n (see equation (126) in the Supplementary Information to Ref. [3]). They then approximate this discrete distribution for large n by a continuous probability distribution density

$$p_t(n) = \frac{1}{\kappa_t} \sqrt{\frac{\mu_t}{n}} \exp\left[-\frac{n + \mu_t}{\kappa_t}\right] I_1\left(\frac{2\sqrt{n\mu_t}}{\kappa_t}\right) \tag{S5}$$

$$\approx \sqrt{\frac{\mu_t^{1/2}}{4\pi\, \kappa_t\, n^{3/2}}} \exp\left[-\frac{(\sqrt{n} - \sqrt{\mu_t})^2}{\kappa_t}\right], \tag{S6}$$

where $I_1(x)$ denotes the modified Bessel function of the first kind. They then use the approximated expression for the analysis of their data (see equation (127) in the

Supplementary Information to Ref. [3])). Equation (S6) depends on two parameters: the mean of lineage size $\mu_t = n_0 e^{\lambda t}$ and the variance to mean ratio $\kappa_t = \frac{1}{2\lambda}(e^{\lambda t} - 1)$, where $\lambda = b - d$ is the mutant growth rate. One important difference between the full solution of the birth-death process and the approximation (S6) is that the full probability distribution has a non-zero weight at $n = 0$ (i.e., extinction is possible) whereas in the approximation (S6) the probability of extinction is zero.

The LB method then uses the same general form of equation (S6) to describe the dynamics of barcode read counts, that is, if a lineage with selection coefficient s is observed to have r_{k-1} read counts at the previous cycle $k-1$, then the probability density $p(r)$ that r counts will be observed at the next cycle is assumed to be

$$p(r; \mu_k, \kappa_k) = \frac{1}{\kappa_t}\sqrt{\frac{\mu_t}{r}} \exp\left[-\frac{r + \mu_t}{\kappa_t}\right] I_1\left(\frac{2\sqrt{r\mu_t}}{\kappa_t}\right) \tag{S7}$$

$$\approx \sqrt{\frac{\mu_k^{1/2}}{4\pi\, \kappa_k\, r^{3/2}}} \exp\left[-\frac{(\sqrt{r} - \sqrt{\mu_k})^2}{\kappa_k}\right], \tag{S8}$$

where

$$\mu_k(s, \bar{s}_k) = \frac{R_k}{R_{k-1}}\, r_{k-1}\, e^{(s-\bar{s}_k)(t_k - t_{k-1})}, \tag{S9}$$

is the expected read count. Here R_{k-1} and R_k are the total read depths at sampling times t_{k-1} and t_k respectively and $\bar{s}_k$ is the mean fitness of the population in the interval (t_{k-1}, t_k). Unlike in the birth-death process, the variance parameter κ_k, which now characterizes all the noise in the BLT experiments including growth, dilution, DNA extraction, DNA amplification, etc., is unknown a priori. To determine κ_k, Levy et al performed additional experiments and estimated $\kappa_k \approx 3$ (see equations (40, 45) in the Supplementary Information to Ref. [3]). However, κ_k is treated as a fitting parameter when BLT data are analyzed, as described below.

Levy et al apply the following procedure to infer the lineage selection coefficients s from the BLT data. At each cycle k, they first infer the mean fitness $\bar{s}_k$ using low-abundance lineages, i.e. those with read counts $20 \leq r_{k-1} \leq 40$. The assumption is that these lineages are so small in size that they have not acquired any adaptive mutations, and so the selection coefficient s for such lineages is zero. For such lineages, equation (S9) simplifies to equation (1) in the main text, which can be used to infer μ_k and therefore $\bar{s}_k$. However, instead, Levy et al construct the entire empirical distribution for r_k and fit equations (S8), (S9) to it, using $\bar{s}_k$ and κ_k as free parameters (see equations (45)–(48) in the Supplementary Information to Ref. [3]).

Having inferred the $\bar{\boldsymbol{s}} = (\bar{s}_1, \bar{s}_2, \dots)$ and $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots)$ from low-abundance lineages, the LB approach takes the entire barcode read-count trajectory $\boldsymbol{r}_i = (r_{i1}, r_{i2}, \dots)$ for each lineage i to infer its selection coefficient s_i and the establishment time τ_i, using a Bayesian approach. Specifically, they define the likelihood

function for the read-count trajectory $\boldsymbol{r}_i$ as

$$L_i(s,\tau) = \prod_k p\big(r_{ik}; \mu_k(s, \bar{s}_k, \tau), \kappa_k\big), \tag{S10}$$

where p is given by equation (S8) and an expression for $\mu_k(s_i, \bar{s}_k, \tau_i)$ that is adjusted to account for the fact that an adapted lineage contains a subpopulation of adapted cells with $s_i > 0$ growing from established time τ_i and a non-adapted subpopulation (see equations (58)–(62) in the Supplementary Information to Ref. [3]). To determine if lineage i is adapted or not, they construct the ratio of posterior probabilities as

$$B_i(s,\tau) = \frac{P^{\text{prior}}(s_i,\tau_i) \times L_i(s_i,\tau_i)}{P^{\text{prior}}(\text{no mutation}) \times L_i(0,0)}, \tag{S11}$$

where P^{prior} denote prior probability for each hypothesis (see equations (64)–(67) in the Supplementary Information to Ref. [3]). Finally, they find values $\hat{s}_i$ and $\hat{\tau}_i$ that maximize B_i. Lineage i is identified as adapted if $B_i(\hat{s}_i, \hat{\tau}_i) > 1$.

2.2 Review of FitMut2

FitMut2 is an improved algorithm based on the LB method. The key difference between the LB method and FitMut2 is that FitMut2 uses an iterative approach to simultaneously call adapted lineages and infer mean fitness from them in a consistent way, which removes the need for a large number of neutral lineages required by the LB method. For convenience, we provide the summary of similarities and differences between all the BLT analysis methods used in this paper in Table S4.

Specifically, for each lineage i, FitMut2 considers two hypotheses, that the lineage acquired an adaptive mutation ($\Theta = 1$) or not ($\Theta = 0$) during the BLT experiment. FitMut2 calculates the probability for the observed read-count trajectory $\boldsymbol{r}_i$ for lineage i as

$$p(\boldsymbol{r}_i) = p(\boldsymbol{r}_i \,|\, \Theta = 0)\, p(\Theta = 0) + p(\boldsymbol{r}_i \,|\, \Theta = 1)\, p(\Theta = 1), \tag{S12}$$

where $p\,(\Theta = 1)$ and $p\,(\Theta = 0)$ are the prior probabilities of the adaptive and neutral hypotheses, respectively.

For the neutral model ($\Theta = 0$), they set $p(\Theta = 0) = 1$ and expand the probability of the trajectory as

$$p(\boldsymbol{r}_i \,|\, \Theta = 0) = p(r_{i0} \,|\, \Theta = 0) \prod_j p(r_{ij} \,|\, r_{ij-1}, \Theta = 0),$$

where $p(r_{i0} \,|\Theta = 0) = p(r_{i0})$ is the initial prior (see equations (14)–(15) in the Supplementary Information to Ref. [4]) and the transition probability $p(r_{ij} \,|\, r_{ij-1}, \Theta = 0)$

is given by equation (S7)) with the expected read count μ_k given by equation (S9) with $s = 0$ (see equation (11) in the Supplementary Information to Ref. [4]).

For the adaptive model ($\Theta = 1$), they express the probability $p\,(\boldsymbol{r}_i \,|\, \Theta = 1)\; p\,(\Theta = 1)$ as an integral over all possible ways to be an adapted lineage with fitness effects s and establishment time τ

$$p(\boldsymbol{r}_i \,|\, \Theta = 1)\, p(\Theta = 1) = \int p(\boldsymbol{r}_i \,|\, s, \tau)\, p(s, \tau)\, ds\, d\tau,$$

where $p(s, \tau)$ is the prior probability that a lineage acquires a beneficial mutation (defined in the Discussion of Ref. [4]). Then the probability of the adaptive trajectory is factorized as

$$p(\boldsymbol{r}_i \,|\, s, \tau) = p(r_{i0} \,|\, s, \tau) \prod_j p(r_{ij} \,|\, r_{ij-1},\, s, \tau),$$

where $p(r_{i0} \,|\, s, \tau) = p(r_{i0})$ is the same initial prior as in the neutral model and $p(r_{ij} \,|\, r_{ij-1},\, s, \tau)$ is given by equation (S7) with the same expected read count μ_k as in the LB method (essentially equation (S9) adjusted for the fact that the adaptive lineage is segregating within the barcoded subpopulation), defined in equation (13) in the Supplementary Information to Ref. [4].

To choose the best model for lineage i, FitMut2 applies Bayes' theorem to calculate the posterior probability of $\Theta = 1$, given the data $\boldsymbol{r}_i$,

$$p(\Theta = 1 \,|\, \boldsymbol{r}_i) = \frac{p(\boldsymbol{r}_i \,|\, \Theta = 1)\, p(\Theta = 1)}{p(\boldsymbol{r}_i)}, \tag{S13}$$

where $p\,(\{r_{ik}\})$ is given by equation (S12). If $p\,(\Theta = 1 \,|\, \boldsymbol{r}_i) > 0.5$, lineage i is called putatively adaptive, with putative estimates of s and τ obtained by maximizing the posterior log-likelihood $L(s, \tau) = \ln\,(p(\boldsymbol{r}_i \,|\, s, \tau)\; p(s, \tau))$.

Then, they calculate the mean fitness trajectory $\bar{\boldsymbol{s}}$ at time k from the frequencies f_{ik} of all putatively adapted lineages i

$$\bar{s}_k = \sum_i f_{ik}\, s_i. \tag{S14}$$

With the newly estimated mean fitness trajectory $\bar{\boldsymbol{s}}$, FitMut2 then re-identifies putatively adaptive lineages and re-estimates their selection coefficients and establishment times. FitMut2 then continues this process iteratively until it converges.

2.3 Key assumptions of the Levy-Blundell method

The LB method—and the neutral decline method as its simplified version—is based on two key assumptions. In this section, we discuss their validity and possible consequences for inference if they are violated.

2.3.1 Assumption 1. Low-abundance lineages are neutral

Low-abundance lineages can be used to infer the mean-fitness trajectory as long as these lineages are neutral, i.e., as long as they have not yet acquired adaptive mutations. If the lineages chosen as reference are in fact not neutral, the inferred mean fitness will be biased, as we demonstrate in the main text. Thus, it is critical for the accuracy of the LB method that this assumption holds.

Of course, as adaptation proceeds, all neutral lineages eventually go extinct and this assumption will eventually be violated. The time scale of persistence of neutral lineages depends on the speed of adaptation, which in turn is determined by the availability of adaptive mutations and their fitness effects, and is therefore a priori unknown.

This assumption can be relaxed if we allow for adapted lineages with known fitness to be used as reference. This can be implemented in practice as follows. By definition, in an initially clonal population, most cells are "wildtype" and any lineage can be used as reference at the initial phase of adaptation. Some lineages will acquire adaptive mutations earlier than others and their fitness can be estimated reliably while neutral lineages are still present in the population. These early adapted lineages can be then used as reference themselves until they acquire secondary adaptive mutations.

2.3.2 Assumption 2. Read frequency dynamics faithfully represent lineage frequency dynamics

According to the population genetics theory, the expected rate of change dx/dt in the frequency of a mutant is determined by its current frequency x, its selection coefficient s and population's mean fitness $\bar{s}$,

$$\frac{dx}{dt} = x(s - \bar{s}).$$

Taking the mean fitness as constant $\bar{s}_k$ over a short period of time $\Delta t_k = t_k - t_{k-1}$ (e.g., between two consecutive sampling time points t_{k-1} and t_k in the BLT experiment), we integrate this equation and obtain the expected frequency x_k at time t_k,

$$x_k = x_{k-1} e^{(s-\bar{s}_k)\Delta t_k}. \tag{S15}$$

Thus, if the lineage frequencies are known, we can solve equation (S15) for $\bar{s}_k$ (if s is known) or for s (if $\bar{s}_k$ is known).

However, x_k are never known with certainty but are estimated. In particular, in BLT experiments, x_k must be estimated from the corresponding barcode read-count data. The LB method (and the neutral decline method as a consequence) makes an implicit but crucial assumption that the dynamics of measured barcode

read frequencies are an unbiased (albeit noisy) representation of the dynamics of *cell* frequencies. While this assumption is reasonable, it is in general not true.

We will first show that this assumption is indeed implied by equation (S9). Since μ_k is the expected number of reads at time t_k, equation (S9) can also be re-written as

$$\frac{\mathbb{E}\left(r_k|r_{k-1}\right)}{R_k} = \frac{r_{k-1}}{R_{k-1}} e^{(s-\bar{s}_k)\Delta t_k}, \tag{S16}$$

which implies that the expected read frequency $\mathbb{E}\left(r_k|r_{k-1}\right)/R_k$ depends on the previous read frequency r_{k-1}/R_{k-1} in exactly the same way as the cell frequency x_k in equation (S15).

We will now show that equation (S16) is in fact not generally true. Suppose that the measurement process is described by the probability distribution $P^{\text{meas}}(r|x;R)$, which determines the number of reads r that are obtained from a lineage whose actual population frequency is x if the total read depth is R. Let us assume that measurement is unbiased, i.e.,

$$\mathbb{E}\left(r|x\right) = \sum_{r=1}^{R} r\, P^{\text{meas}}(r|x;R) = xR, \tag{S17}$$

which is supported by our data (see Figure 4A in the main text). We rewrite $\mathbb{E}\left(r_k|r_{k-1}\right)$ in terms of the measurement process P^{meas} and the evolutionary process P^{evol}.

$$\begin{aligned}
\mathbb{E}\left(r_k|r_{k-1}\right) &= \int_0^1 dx_{k-1} \int_0^1 dx_k\, P^{\text{evol}}(x_k|x_{k-1})\, P(x_{k-1}|r_{k-1}) \sum_{r_k} r_k P^{\text{meas}}(r_k|x_k) \\
&= R_k \int_0^1 dx_{k-1}\, P(x_{k-1}|r_{k-1}) \int_0^1 dx_k\, x_k\, P^{\text{evol}}(x_k|x_{k-1}) \\
&= R_k\, e^{(s-\bar{s}_k)\Delta t_k} \int_0^1 dx_{k-1}\, x_{k-1}\, P(x_{k-1}|r_{k-1}) \\
&= R_k\, e^{(s-\bar{s}_k)\Delta t_k}\, \mathbb{E}\left(x_{k-1}|r_{k-1}\right),
\end{aligned} \tag{S18}$$

where to simplify the notations we dropped the parameters from the expressions for various probability distributions. Equation (S18) shows that equation (S16) would be true if the expected frequency of lineages represented by r out of R reads is in fact r/R,

$$\mathbb{E}\left(x|r\right) = \frac{r}{R}. \tag{S19}$$

Counterintuitively, equation (S19) is in general not true. To demonstrate this, we first use the Bayes' theorem to re-write it as

$$\mathbb{E}\left(x|r\right) = \int_0^1 x\, P(x|r)\, dx = \int_0^1 x\, \frac{P^{\text{meas}}(r|x)\, P^{\text{prior}}(x)}{P(r)}\, dx. \tag{S20}$$

Here, $P(r)$ is the normalization constant and $P^{\text{prior}}(x)$ is the a priori probability that the frequency of the focal lineage in the population is x. Note that in general $P^{\text{prior}}(x)$ is not the distribution of lineage frequencies in the population. Indeed, at later time points in the BLT experiment, $P^{\text{prior}}(x)$ for different lineages are different, because they are informed by previous observations. However, at the beginning of the BLT experiment, all lineages are a priori identical, and $P^{\text{prior}}(x)$ can be approximated by the distribution of lineage frequencies in the population. Previous BLT studies have shown that, despite efforts to introduce all barcodes at the same frequency, the initial distributions of barcode frequencies are often be quite broad and may be close to exponential [2]. We now use four specific examples to show that equation (S19) does not hold.

Example 1. Poisson measurement noise and a uniform frequency distribution. First, we consider the simplest case, when the prior frequency distribution is uniform and the measurement noise is Poisson, i.e.,

$$\begin{aligned} P^{\text{prior}}(x)\,dx &= dx, \\ P^{\text{meas}}(r|x;R) &= \frac{(xR)^r\,e^{-xR}}{r!}. \end{aligned} \tag{S21}$$

We show in Appendix B, that the conditional distribution for frequency x given the read count r is approximately a gamma distribution with shape parameter $r+1$ and scale parameter $1/R$, such that

$$\mathbb{E}\left(x|r\right) \approx \frac{r+1}{R}. \tag{S22}$$

Reassuringly, when $r \gg 1$, expression (S22) converges to r/R, suggesting that Assumption 2 is reasonably accurate as long as all lineages are represented by many reads. However, at small numbers of reads, expression (S22) can deviate substantially from equation (S19).

Example 2. Poisson measurement noise and an exponential frequency distribution. We now consider an exponential frequency distribution, which may be adequate for initial time points [2],

$$P^{\text{prior}}(x)\,dx = N_L e^{-N_L x}\,dx, \tag{S23}$$

where N_L is the number of barcoded lineages, such that $\mathbb{E}\left(x\right) = 1/N_L$. We show in Appendix B, that the conditional distribution for frequency x given the read count r is approximately a gamma distribution with shape parameter $r+1$ and scale parameter $1/(R+N_L)$, such that

$$\mathbb{E}\left(x|r\right) \approx \frac{r+1}{R+N_L}. \tag{S24}$$

Since $R/N_L \gtrsim 100$ in typical BLT experiments, expression (S24) is still close to r/R for large r, but it again can deviate substantially from (S19) when the number of reads is small.

Example 3. Measurement noise with an increasing variance to mean ratio and a Gamma frequency distribution. We now consider a more realistic noise model, where the variance to the mean ratio increases with the mean according to equation (2) in the main text. We model the read count as a negative binomial random variable with mean xR and the variance to the mean ratio $1 + \epsilon xR$.

$$P^{\text{meas}}(r|x; R, \epsilon) = \frac{\Gamma(r + 1/\epsilon)}{\Gamma(r + 1)\,\Gamma\,(1/\epsilon)}(1 - p)^r\, p^{1/\epsilon}. \tag{S25}$$

with $p = (1+\epsilon xR)^{-1}$. We also consider a Gamma frequency distribution with shape parameter α and scale parameter $1/\tilde{N}$,

$$P^{\text{prior}}(x) = \frac{\tilde{N}^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\tilde{N}x}, \tag{S26}$$

which may be a more accurate description of our prior knowledge at later time points with $\alpha/\tilde{N}$ representing the a priori expected frequency. If $\epsilon \ll 1$ and $r \ll 1/\epsilon$, the posterior distribution is approximately Gamma with shape parameter $r + \alpha$ and scale parameter $(R + \tilde{N})^{-1}$ (see Appendix B), such that

$$\mathbb{E}\,(x|r) = \frac{r + \alpha}{R + \tilde{N}}. \tag{S27}$$

Example 4. Measurement noise with a constant variance to mean ratio and an exponential frequency distribution. Finally, we consider a noise model where the variance to the mean ratio is constant and equal to 2κ as in Ref. [3]. We again model the read count as a negative binomial random variable,

$$P^{\text{meas}}(r|x) = \frac{\Gamma(r + k)}{\Gamma(r + 1)\Gamma(k)}(1 - p)^r p^k.$$

where $p = 1/2\kappa$ and $k = xR/(2\kappa - 1)$. In this case, as we show in Appendix B, we have

$$\mathbb{E}\,(x|r) \approx \frac{2\kappa + r}{R}, \quad r = 0, 1, \ldots, R. \tag{S28}$$

Substituting any of the equations (S22), (S24), (S27) or (S28) into equation (S18), we see that the relationship between the expected number of reads at time t_k and the observed number of reads at t_{k-1} is not the same as the relationship (S15) between the expected lineage frequency at time t_k and the actual frequency at t_{k-1}.

Specifically, unlike the relationship between $\mathbb{E}(x_k|x_{k-1})$ and x_{k-1}, the relationship between $\mathbb{E}(r_k|r_{k-1})$ and r_{k-1} appears to generally have a non-zero y-intercept, whose value depends on the parameters of the prior distribution and measurement noise. However, these two relationships become identical when r_{k-1} is sufficiently large, although how large it needs to be depends on the noise model.

3 BASIL

3.1 Model of measurement

As discussed in the main text, we cannot directly observe the size n of a barcoded lineage. Instead, a sample from the population is sequenced, and we observe the number of reads that contain the focal barcode.

We model the measurement process with the negative binomial distribution with mean $\langle r \rangle = nR/N$ and variance $\sigma_r^2 = \langle r \rangle + \epsilon \langle r \rangle^2$ (see Section 1), where N is total number of cells in the population during sampling, R is the total coverage, and r is the number of reads that contain the focal barcode, and ϵ is a free parameter controlling overdispersion. The negative binomial distribution is typically parameterized by the number of "successes" k and the success probability p, with mean and variance expressed in terms of these parameters as

$$\begin{aligned} \langle r \rangle &= \frac{k(1-p)}{p} \\ \sigma_r^2 &= \frac{k(1-p)}{p^2} = \frac{\langle r \rangle}{p}, \end{aligned}$$

which yield

$$\begin{aligned} k &= \frac{p}{1-p} \langle r \rangle = \frac{1}{\epsilon}, \\ p &= \frac{1}{1+\epsilon \langle r \rangle}. \end{aligned} \tag{S29}$$

Thus, the probability of observing r reads for lineage with n cells is

$$P^{\text{meas}}(r|n;\epsilon) = \frac{\Gamma(r+1/\epsilon)}{\Gamma(r+1)\Gamma(1/\epsilon)} \left(\frac{\epsilon \langle r \rangle}{1+\epsilon \langle r \rangle}\right)^r \left(\frac{1}{1+\epsilon \langle r \rangle}\right)^{\frac{1}{\epsilon}} \tag{S30}$$

with $\langle r \rangle = nR/N$. The measurement distribution (S30) depends on the noise parameter ϵ, which we fit, as well as on N and R which are assumed to be known. Furthermore, this distribution becomes a point measure at $r = 0$ if $n = 0$ and it converges to the Poisson distribution with mean $\langle r \rangle = nR/N$ as $\epsilon \to 0$.

3.2 The belief distribution for lineage size and fitness

Our main object is the belief probability distribution $P^{\text{belief}}(n, s \,|\, \boldsymbol{r})$ that a lineage with the observation vector $\boldsymbol{r} = (r_0, \ldots, r_k)$ has size n and selection coefficient s. For a given lineage i whose observation vector up to and including time point t_k is $\boldsymbol{r}_{ik} = (r_{i0}, \ldots, r_{ik})$, we will sometimes use a simplified notation $P^{\text{belief}}_{ik}(n, s)$ instead of $P^{\text{belief}}(n, s \,|\, \boldsymbol{r}_{ik})$. For mathematical tractability, we express the belief distribution as

$$P^{\text{belief}}(n, s|\boldsymbol{r}) = P^{\text{belief}}(n|s, \boldsymbol{r})\, P^{\text{belief}}(s|\boldsymbol{r}) \tag{S31}$$

and assume the following parametric forms

$$P^{\text{belief}}(n|s, \boldsymbol{r}) \;= \gamma\left(n\,;\, \varkappa(s, \boldsymbol{r}), \theta(s, \boldsymbol{r})\right), \tag{S32}$$

$$P^{\text{belief}}(s|\boldsymbol{r}) = N\left(s\,;\, \mu(\boldsymbol{r}), \sigma^2(\boldsymbol{r})\right), \tag{S33}$$

where $N(s; \mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2 and $\gamma\,(n; \varkappa, \theta)$ is the gamma distribution with shape parameter $\varkappa$ and scale parameter θ.

Thus, $P^{\text{belief}}(n, s|\boldsymbol{r})$ belongs to a family of distributions with four parameters μ, σ^2, $\varkappa$, θ where the latter two parameters can be functions of s. In this section, we describe how we update these parameters as the observation vector is augmented from one time point to the next. We do so in two steps, which we refer to as "projection" and "update". During the projection step, which is described in Section 3.2.1, we project the past belief distribution P^{belief}_{ik-1} to obtain the prior distribution P^{prior}_{ik} for the time point t_k. During the update step, which is described in Section 3.2.2, we use the Bayes' theorem to obtain the new belief distribution P^{belief}_{ik} based on the prior distribution P^{prior}_{ik} and the read count r_{ik} observed at time t_k (see equation (3) in the main text). Importantly, the parametric distribution given by equation (S32) has no weight at $n = 0$. In other words, it is valid only for lineages that are present in the population and have not yet gone extinct. We describe how we treat lineage extinctions in Section 3.2.2.

3.2.1 Projecting the belief distribution

To obtain the prior distribution $P^{\text{prior}}(n, s|\boldsymbol{r}')$ for the time point t_k, where $\boldsymbol{r}' = (r_0, \ldots, r_{k-1})$ is the observation vector up to and including the previous time point t_{k-1}, we assume that the belief distribution $P^{\text{belief}}(n, s|\boldsymbol{r}')$ at the previous time point is known and has the parametric form given by equations (S32), (S33) with parameters μ, σ, θ, $\varkappa$. To simplify notations, in this section, we will omit the explicit dependence of all probabilities on the observation vector $\boldsymbol{r}'$.

The sampling time points $t_0, t_1, \ldots$ in the BLT experiment may be separated by one or multiple growth and dilution cycles. We will first consider one such cycle and

obtain the "1-cycle projected distribution" $P_1(n,s)$ whose parameters we denote as μ_1, σ_1, $\varkappa_1$, and θ_1. We then use $P_1(n,s)$ to obtain the prior probability for n and s after multiple cycles.

To derive $P_1(n,s)$, we consider a barcoded lineage that has fitness s relative to the ancestor and that is represented by n cells immediately prior to dilution. At dilution, a fraction $1/D$ of the population is transferred into fresh medium, and the rest is discarded. Immediately after dilution, the lineage's size becomes n', which we model as a Poisson random variable with mean n/D, such that

$$P^{\text{dil}}(n') = \frac{\left(\frac{n}{D}\right)^{n'} e^{-n/D}}{\Gamma(n'+1)}, \quad n' = 0, 1, \ldots$$

We assume that after dilution, the lineage grows deterministically based on the difference between its fitness s and the population's mean fitness $\bar{s}$, such that by the end of the cycle, its size is $n_1 = An'$, where

$$A = De^{(s-\bar{s})\Delta t_c}$$

and Δt_c is the cycle duration in generations. Then, the one-cycle transition probability is

$$P^{\text{c}}(n_1 \,|\, n; s, \bar{s}) = P^{\text{dil}}\left(\frac{n_1}{A}\right) = \frac{\left(\frac{n}{D}\right)^{n_1/A} e^{-n/D}}{\Gamma\left(\frac{n_1}{A}+1\right)}, \tag{S34}$$

where formally $n_1/A = 0, 1, 2, \ldots$. Then, for the 1-cycle projected distribution we have

$$P_1(n_1, s) = \int_0^\infty P^{\text{c}}(n_1 \,|\, n; s, \bar{s})\; P^{\text{belief}}(n, s)\; dn. \tag{S35}$$

Integrating equation (S35) with respect to n_1, we find, as expected, that the projection does not change the marginal distribution for s,

$$P_1(s) = P^{\text{belief}}(s) = N\left(s\,;\,\mu, \sigma^2\right),$$

which implies that $\mu_1 = \mu$ and $\sigma_1 = \sigma$. Since we can express the joint distribution $P_1(n_1, s)$ through this marginal distribution as $P_1(n_1, s) = P_1(n_1 \,|\, s)\; P_1(s)$, we now only need to obtain the conditional probability $P_1(n_1 \,|\, s)$. To do so, we note that

$$\begin{aligned} P_1(n_1 \,|\, s) &= \frac{P_1(n_1, s)}{P_1(s)} = \frac{\int_0^\infty dn\, P^{\text{c}}(n_1 \,|\, n; s, \bar{s}_k)\; P^{\text{belief}}(n \,|\, s)\, P^{\text{belief}}(s)}{P^{\text{belief}}(s)} \\ &= \int_0^\infty P^{\text{c}}(n_1 \,|\, n; s, \bar{s}_k)\; P^{\text{belief}}(n \,|\, s)\, dn. \end{aligned} \tag{S36}$$

Substituting expressions (S34) and (S32) into equation (S36), we obtain

$$
\begin{aligned}
P_1\left(n_1 \,|\, s\right) &= \frac{1}{D^{\frac{n_1}{A}}\,\Gamma\left(\frac{n_1}{A}+1\right)\,\Gamma(\varkappa)\theta^{\varkappa}} \int_0^{\infty} n^{\tilde{\varkappa}-1} e^{-\frac{n}{\tilde{\theta}}}\, dn \\
&= \frac{\Gamma\left(\frac{n_1}{A}+\varkappa\right)}{\Gamma\left(\frac{n_1}{A}+1\right)\,\Gamma(\varkappa)} \left(\frac{\theta}{D+\theta}\right)^{\frac{n_1}{A}} \left(\frac{D}{D+\theta}\right)^{\varkappa},
\end{aligned}
\tag{S37}
$$

with $n_1/A = 0, 1, 2, \ldots$, $\tilde{\varkappa} = \frac{n_1}{A} + \varkappa$, $\tilde{\theta} = \frac{\theta D}{\theta + D}$. Equation (S37) shows that, conditional on s, n_1/A is distributed as negative binomial with success probability $D/(D+\theta)$ and number of successes $\varkappa$. Therefore,

$$
\begin{aligned}
\mathbb{E}\left(n_1|s\right) &= \varkappa\,\theta\, e^{(s-\bar{s})\Delta t_c}, \\
\mathrm{Var}\left(n_1|s\right) &= \varkappa\,\theta\,(D+\theta) e^{2(s-\bar{s})\Delta t_c}.
\end{aligned}
$$

Furthermore, the lineage goes extinct after one growth and dilution cycle with probability

$$
p_{\mathrm{ext},1}(\theta, \varkappa) = \left(\frac{D}{D+\theta}\right)^{\varkappa} \tag{S38}
$$

and survives with probability

$$
p_{\mathrm{surv},1}(\theta, \varkappa) = 1 - \left(\frac{D}{D+\theta}\right)^{\varkappa}. \tag{S39}
$$

It is then easy to show that the mean and variance conditional on survival are

$$
\mathbb{E}\left(n_1 \,|\, s, \mathrm{surv}\right) = \frac{\varkappa\,\theta}{p_{\mathrm{surv},1}(\theta, \varkappa)}\, e^{(s-\bar{s})\Delta t_c}, \tag{S40}
$$

$$
\mathrm{Var}\left(n_1 \,|\, s, \mathrm{surv}\right) = \frac{p_{\mathrm{surv},1}(\theta, \varkappa)\,(D+\theta+\varkappa\,\theta) - \varkappa\,\theta}{p^2_{\mathrm{surv},1}(\theta, \varkappa)} \varkappa\,\theta\, e^{2(s-\bar{s})\Delta t_c}. \tag{S41}
$$

To facilitate further calculations, we approximate the distribution of n_1, conditional on s and lineage survival, by a gamma distribution with the shape parameter $\varkappa_1$ and scale parameter θ_1. To determine $\varkappa_1$ and θ_1, we equate the mean $\varkappa_1\theta_1$ and variance $\varkappa_1\theta_1^2$ of this gamma distribution to the conditional mean and variances given by equations (S40) and (S41) and find $\theta_1 = f_\theta(s, \theta, \varkappa)$ and $\varkappa_1 = f_\varkappa(\theta, \varkappa)$ where

$$
f_\theta(s, \theta, \varkappa) = \frac{p_{\mathrm{surv},1}(\theta, \varkappa)\,(\theta + D + \varkappa\theta) - \varkappa\theta}{p_{\mathrm{surv},1}(\theta, \varkappa)}\, e^{(s-\bar{s}_k)\Delta t_c}, \tag{S42}
$$

$$
f_\varkappa(\theta, \varkappa) = \frac{\varkappa\theta}{p_{\mathrm{surv},1}(\theta, \varkappa)\,(\theta + D + \varkappa\theta) - \varkappa\theta}. \tag{S43}
$$

Therefore, the full projected probability for a lineage after one growth and dilution cycle is

$$P_1(n,s) = N\left(s;\mu,\sigma\right) \times \begin{cases} p_{\text{ext},1}(\theta,\varkappa)\,, & \text{if } n=0, \\ p_{\text{surv},1}(\theta,\varkappa)\,\gamma\left(n;\varkappa_1,\theta_1\right), & \text{if } n>0, \end{cases}$$

where $p_{\text{ext},1}$ and $p_{\text{surv},1}$ are given by equations (S38) and (S39) and parameters θ_1 and $\varkappa_1$ are the functions of θ and $\varkappa$ given by equations (S42) and (S43).

If the next sampling occurs at the end of a single growth and dilution cycle after the previous sampling, then we set $P^{\text{prior}}(n,s) = P_1(n,s)$. If $L > 1$ growth and dilution cycles elapse between successive samples, then we set the prior probability $P^{\text{prior}}(n,s)$ for the next sampling time point to be equal to

$$P^{\text{prior}}(n,s) = N\left(s;\mu,\sigma\right) \times \begin{cases} 1 - p_{\text{surv},L}\,, & \text{if } n=0, \\ p_{\text{surv},L}\,\gamma\left(n;\varkappa_L,\theta_L\right), & \text{if } n>0, \end{cases} \tag{S44}$$

where the L-cycle survival probability is given by

$$p_{\text{surv},L} = \prod_{\ell=1}^{L} p_{\text{surv},1}(\theta_\ell,\varkappa_\ell), \tag{S45}$$

and parameters $\theta_\ell = f_\theta\left(s,\theta_{\ell-1},\varkappa_{\ell-1}\right)$, $\varkappa_\ell = f_\varkappa\left(\theta_{\ell-1},\varkappa_{\ell-1}\right)$ for $\ell = 1,2,\ldots L$ are obtained recursively using equations (S42) and (S43) with $\varkappa_0 \equiv \varkappa$, $\theta_0 \equiv \theta$. Note that since all $\varkappa_\ell$ and θ_ℓ depend on the parameters $\varkappa$, θ of the prior belief distribution $P^{\text{belief}}(n\,|\,s,\boldsymbol{r}')$, which themselves are functions of s, the survival probability $p_{\text{surv},L}$ is also a function of s.

3.2.2 Updating the belief distribution

Next, we apply the Bayes' theorem to obtain the belief distribution $P^{\text{belief}}\left(n,s\,|\,\boldsymbol{r}\right)$ after observing the current read count r_k,

$$P^{\text{belief}}\left(n,s\,|\,\boldsymbol{r}\right) = \frac{P^{\text{prior}}\left(n,s\,|\,\boldsymbol{r}'\right)\,P^{\text{meas}}\left(r_k\,|\,n\right)}{Z^{\text{sel}}(\boldsymbol{r})}, \tag{S46}$$

where $\boldsymbol{r}' = (r_0,\ldots,r_{k-1})$ and $\boldsymbol{r} = (r_0,\ldots,r_{k-1},r_k)$ are the previous and current observation vectors, $P^{\text{prior}}\left(n,s\,|\,\boldsymbol{r}'\right)$ is the prior probability that the lineage has selection coefficient s and size n at the current observation time point before the current measurement is made (see equation (S44)), $P^{\text{meas}}\left(r\,|\,n\right)$ is the probability of observing r reads for a lineage of size n (see equation (S30)), and the denominator

$$Z^{\text{sel}}(\boldsymbol{r}) = \int_{-\infty}^{\infty} ds \int_0^{\infty} dn\, P^{\text{prior}}\left(n,s\,|\,\boldsymbol{r}'\right)\,P^{\text{meas}}\left(r_k\,|\,n\right). \tag{S47}$$

is the marginal probability of observing r_k reads in our model. The superscript "sel" stands for "selection" and denotes the fact that our model allows the selection coefficient of the lineage to vary freely. We will later consider a neutral null model where the selection coefficient is fixed at zero (see Section 3.3). Note that P^{prior} depends on $\bar{s}_k$, the mean fitness at the current interval (t_{k-1}, t_k), and P^{meas} depends on the noise parameters ϵ_k at the current interval (t_{k-1}, t_k). Therefore, both Z^{sel} and P^{belief} depend on both $\bar{s}_k$ and ϵ_k. For now, we can ignore these dependencies treating these parameters as fixed and known. However, these dependencies will becomes important in Section 3.3, where we discuss how $\bar{s}_k$ and ϵ_k are estimated.

As discussed above, we would like to approximate the belief distribution with an analytical parametric form given by equations (S32), (S33). However, since the numerator of equation (S46) is a complex function of both n and s, the normalization constant (S47) and the moments of this distribution cannot be expressed analytically. Therefore, to estimate the parameters of the belief distribution, we draw a random sample (n_j, s_j), $j = 1, 2, \ldots, M$ from the belief distribution $P^{\text{belief}}(n, s \,|\, \boldsymbol{r})$ using the Markov Chain Monte Carlo (MCMC) approach (see Section 3.5 for the details of the algorithm). We consider two cases, $r_k > 0$ and $r_k = 0$, which differ qualitatively because in the former case, the lineage is guaranteed to have survived until the current sampling time point t_k whereas in the later case it is possible that the lineage has gone extinct.

Guaranteed lineage survival when $r_k > 0$. If $r_k > 0$, that is, if the lineage is observed at current time point t_k, then $P^{\text{belief}}(0, s \,|\, \boldsymbol{r}) = 0$ since $P^{\text{meas}}(r_k \,|\, 0) = 0$ whenever $r_k > 0$. In other words, if we observe the lineage, we know with certainty that it has survived up to the current sampling time point. We then estimate the parameters μ, σ of the distribution (S33) from this sample as

$$\hat{\mu} = \frac{1}{M} \sum_{j=1}^{M} s_j, \tag{S48}$$

$$\hat{\sigma} = \sqrt{\frac{1}{M-1} \sum_{j=1}^{M} (s_j - \hat{\mu})^2}. \tag{S49}$$

Estimating parameters $\varkappa(s)$ and $\theta(s)$ of the conditional distribution (S32) is more difficult because they can be arbitrary functions of s. In principle, one could estimate these functions by independently fitting the shape and scale parameters of the gamma distribution for each s-slice of the joint distribution. However, this approach is computationally intensive. To reduce computational burden, we sought to find simple functional forms for the functions $\varkappa(s)$ and $\theta(s)$ and fitting the parameters of these functional forms directly from the joint distribution. To this end, we no-

ticed that $\ln n$ and s are linearly correlated in our MCMC samples (see Figure S3), implying that

$$\mathbb{E}\left(\ln n|s\right) \approx \tilde{a} + \tilde{b}\, s. \tag{S50}$$

Since

$$\mathbb{E}\left(\ln n|s\right) = \psi(\varkappa) + \ln\theta \approx \ln\varkappa + \ln\theta,$$

where ψ is the di-gamma function (see Appendix C) and the approximation $\psi(\varkappa) \approx \ln\varkappa$ holds when $\varkappa \gg 1$, we obtain

$$\theta \approx \frac{e^{\tilde{a}+\tilde{b}s}}{\varkappa},$$

which suggests that an exponential dependence of θ on s should capture the shape of our belief distribution. To simplify subsequent calculations, we use the ansatz

$$\theta\left(s\right) = \frac{a}{\varkappa}\exp\left[\frac{b\left(s-\mu\right)}{\sigma}\right], \tag{S51}$$

where $\varkappa > 0$, a and b are the new real-valued parameters of distribution (S32).

With this parametrization, we have

$$\mathbb{E}\left(n\right) = a\, e^{\frac{b^2}{2}}, \tag{S52}$$

$$\mathrm{Var}\left(n\right) = \frac{a^2 e^{b^2}}{\varkappa}\left((1+\varkappa)e^{b^2} - \varkappa\right), \tag{S53}$$

$$\mathrm{Cov}\left(n, s\right) = ab\sigma\, e^{\frac{b^2}{2}}. \tag{S54}$$

Equations (S52)–(S54) can be solved to yield estimators of the parameters of the gamma distribution (S32) with the functional form (S51),

$$\hat{b} = \frac{\widehat{\mathrm{Cov}}\left(s, n\right)}{\langle n\rangle\ \hat{\sigma}} \tag{S55}$$

$$\hat{a} = \langle n\rangle\, e^{-\hat{b}^2/2} \tag{S56}$$

$$\hat{\varkappa} = \frac{\hat{a}^2 e^{2\hat{b}^2}}{\widehat{\mathrm{Var}}(n) - \hat{a}^2 e^{\hat{b}^2}(e^{\hat{b}^2} - 1)}, \tag{S57}$$

where

$$\langle n \rangle = \frac{1}{M} \sum_{j=1}^{M} n_j, \tag{S58}$$

$$\widehat{\mathrm{Var}}(n) = \frac{1}{M-1} \sum_{j=1}^{M} (n_j - \langle n \rangle)^2, \tag{S59}$$

$$\widehat{\mathrm{Cov}}(s, n) = \frac{1}{M-1} \sum_{j=1}^{M} (s_j - \hat{\mu})(n_j - \langle n \rangle). \tag{S60}$$

To test how well our parametric form (S32), (S33), (S51) with the parameters fitted using equations (S55)–(S60) fits the MCMC-sampled posterior distribution, we selected four lineages from simulated and real data and then plotted their MCMC-sampled marginal distributions for s and n as well as their joint distribution (Figure S3). To calculate the marginal distribution for n, we generated $M' = 10^5$ random samples of s_j, $j = 1, \ldots, M'$ from the normal distribution with mean $\hat{\mu}$ and variance $\hat{\sigma}^2$. We then approximate the marginal distribution for n as

$$P^{\mathrm{belief}}(n|\boldsymbol{r}) \approx \frac{1}{M'} \sum_{j=1}^{M'} \gamma\left(n\,;\, \hat{\varkappa}, \hat{\theta}_j\right), \tag{S61}$$

where $\hat{\theta}_j = \hat{a}/\hat{\varkappa} \exp\left(\hat{b}\,(s_j - \hat{\mu})/\hat{\sigma}\right)$ with s. Figure S3 shows that our parametric form captures the MCMC-sampled distribution reasonably well.

Possible lineage extinction when $r_k = 0$. If $r_k = 0$, there are two possibilities. With probability $p_{\mathrm{surv},L}$ given by equation (S45), the lineage has survived up to the current sampling time point t_k but was not detected due to sampling noise; or, with probability $1 - p_{\mathrm{surv},L}$, the lineage has gone extinct between t_{k-1} and t_k. Therefore, the belief distribution $P^{\mathrm{belief}}(n, s|\boldsymbol{r})$ has a non-zero weight at $n = 0$,

$$P^{\mathrm{belief}}(n, s) = \frac{N(s; \mu, \sigma)}{Z^{\mathrm{sel}}} \times \begin{cases} 1 - p_{\mathrm{surv},L}\,, & \text{if } n = 0, \\ p_{\mathrm{surv},L}\, \frac{\gamma(n; \varkappa_L, \theta_L)}{(1+\epsilon_k n R/N)^{1/\epsilon_k}}, & \text{if } n > 0, \end{cases} \tag{S62}$$

with

$$\begin{aligned} Z^{\mathrm{sel}} &= \int_{-\infty}^{\infty} N(s; \mu, \sigma) \left(1 - p_{\mathrm{surv},L} + p_{\mathrm{surv},L} \int_0^{\infty} \frac{\gamma(n; \varkappa_L, \theta_L)\, dn}{(1 + \epsilon_k n R/N)^{1/\epsilon_k}}, \right) ds \\ &\approx \int_{-\infty}^{\infty} N(s; \mu, \sigma) \left(1 - p_{\mathrm{surv},L} + p_{\mathrm{surv},L} \left(\frac{R}{N\theta_L + R}\right)^{\varkappa_L}\right) ds, \end{aligned}$$

where the approximation holds when $\epsilon_k \ll 1$, and we have omitted the dependencies on the observation vector $\boldsymbol{r}$ and mean fitness $\bar{s}_k$.

Equation (S62) shows that the belief distribution can no longer be adequately captured by the parametric form given by equations (S32), (S33). Instead, we use this parametric form to capture the conditional belief distribution, conditional on the lineage having survived until the current sampling time point t_k. We do so using the same approach as above, i.e., we obtain an MCMC sample (n_j, s_j), $j = 1, 2, \ldots, M$ from the distribution (S62) and then use equations (S58)–(S60) to estimate the mean and the variance of the lineage size n and the covariance between lineage size and selection coefficient, except we use only those MCMC samples with $n_j > 0$. We then apply equations (S48), (S49), (S55)–(S57) to estimate the parameters of the conditional belief distribution (S32), (S33). Furthermore, if the lineage was not observed at two consecutive time points, we assume that it has gone extinct and stop tracking it.

3.3 Estimating populations mean fitness and the noise parameter

As mentioned above, our model of evolution and consequently our procedure for updating the belief distribution at the "current" time point t_k depends on the population's mean fitness $\bar{s}_k$ during the interval (t_{k-1}, t_k), $k \geq 1$ as a parameter (see equation (S34)). Similarly, our measurement distribution (equation (S30)) depends on the noise parameter ϵ_k during the interval (t_{k-1}, t_k). So far, we have assumed that both $\bar{s}_k$ and ϵ_k are fixed and known. However, in reality the true values of $\bar{s}_k$ and ϵ_k are of course unknown and must be estimated before we update the lineage belief distributions at t_k.

We estimate $\bar{s}_k$ and ϵ_k using the maximum likelihood approach. To this end, we choose 3000 random lineages for which $r_{ik-1} > 0$ and classify each of them as either putatively neutral or putatively adapted (at the current time interval (t_{k-1}, t_k)) as described below. We denote the subsets of currently putatively neutral and putatively adapted lineages as $\mathcal{N}_k$ and $\mathcal{A}_k$, respectively. For each putatively adapted lineage $i \in \mathcal{A}_k$, we can treat the normalization constant $Z^{\text{sel}}(\boldsymbol{r}_{ik}; \bar{s}_k, \epsilon_k)$ given by equation (S47) as the likelihood of observing the number of reads r_{ik} for that lineage at time t_k under the model with selection. Similarly, for each putatively neutral lineage $i \in \mathcal{N}_k$, we can treat the normalization constant $Z^{\text{neut}}(\boldsymbol{r}_{ik}; \bar{s}_k, \epsilon_k)$ under the neutral model described below (see equation (S66)) as the likelihood of observing the number of reads r_{ik} for that lineage at time t_k under the neutral model. The procedures for estimating $Z^{\text{sel}}(\boldsymbol{r}_{ik}; \bar{s}_k, \epsilon_k)$ and $Z^{\text{neut}}(\boldsymbol{r}_{ik}; \bar{s}_k, \epsilon_k)$ are described below. Since both $Z^{\text{sel}}(\boldsymbol{r}_{ik}; \bar{s}_k, \epsilon_k)$ and $Z^{\text{neut}}(\boldsymbol{r}_{ik}; \bar{s}_k, \epsilon_k)$ depend on the unknown parameters

$\bar{s}_k$ and ϵ_k, we can write down the log-likelihood function for $\bar{s}_k$ and ϵ_k as

$$L_k(\bar{s}, \epsilon) = \sum_{i \in \mathcal{A}_k} \log Z^{\text{sel}}\left(\boldsymbol{r}_{ik}; \bar{s}, \epsilon\right) + \sum_{i \in \mathcal{N}_k} \log Z^{\text{neut}}\left(\boldsymbol{r}_{ik}; \bar{s}, \epsilon\right), \tag{S63}$$

and estimate

$$(\bar{s}_k, \epsilon_k) = \arg\max_{\bar{s}, \epsilon} L_k(\bar{s}, \epsilon)$$

using standard optimization algorithms.

Classification of lineages for estimating mean fitness. We classify lineages into putatively neutral and putatively adapted to minimize estimation biases that may arise and be amplified over time due to the following positive feedback loop. Evolutionary dynamics in our model depend only on the difference between the fitness of a lineage and the mean fitness (see equations (S42),(S43)), but not on these two quantities individually. Thus, all selection coefficients and the mean fitness can be identified from data only up to a shared constant. We eliminate this ambiguity by assuming that the expected fitness of all lineages is zero at t_0 (see Section **??**). In other words, we estimate selection coefficients with respect to population's initial mean fitness. While this condition should be sufficient to accurately estimate the mean fitness and lineage selection coefficients on average, the large uncertainty in our priors can lead to large uncertainties in the initial estimates of lineage fitness (Figure S4) and mean fitness. These initial deviations can over time be "baked in" into the belief distributions because, as mentioned above, additional data can only correct any inaccuracies in the difference between lineage fitness and mean fitness but not in their individual values.

To make this reasoning more concrete, suppose that we do not classify lineages into putatively neutral and adapted and instead use lineage belief distributions to estimate mean fitness, i.e., use only Z^{sel} terms for all lineages in equation (S63). Imagine that to estimate mean fitness $\bar{s}_1$ at the first time interval (t_0, t_1), by chance we pick 3,000 lineages whose frequency declines (slightly) more than expected under neutrality. Then, we will (slightly) over-estimate $\bar{s}_1$. As a result, the projected distributions would under-estimate lineage sizes at t_1, thus generating an overly strong surprise once the read counts are observed. This would push us to update the belief distribution towards overly high s for many lineages, which would in turn cause an additional over-estimate of the mean fitness at the next time interval (t_1, t_2), etc.

To dampen this positive feedback loop at the mean-fitness estimation step, we assume that all lineages are neutral by default (for the purposes of mean-fitness estimation) unless we have a high degree of confidence that a given lineage is adapted. Specifically, we classify lineage i as putatively adapted for the purposes of estimating

$\bar{s}_k$ if

$$\mu_{ik-1} > 3\,\sigma_{ik-1}, \tag{S64}$$

where

$$\mu_{ik-1} = \int_{-\infty}^{\infty} ds \int_{0}^{\infty} s\, P^{\text{belief}}\left(n, s \,|\, \boldsymbol{r}_{ik-1}\right)\, dn,$$
$$\sigma_{ik-1}^2 = \int_{-\infty}^{\infty} ds \int_{0}^{\infty} \left(s - \mu_{ik-1}\right)^2\, P^{\text{belief}}\left(n, s \,|\, \boldsymbol{r}_{ik-1}\right)\, dn$$

are the mean and variance of the marginal belief distribution for s for lineage i at time t_{k-1} estimated using equations (S48), (S49). The factor 3 in equation (S64) was chosen so that our confidence that the lineage is in fact adapted (i.e., has a positive s) exceeds 99%.

Neutral model of lineage evolution. To compute the marginal likelihood $Z^{\text{neut}}(\boldsymbol{r}_{ik}; \bar{s}_k, \epsilon_k)$ that r_{ik} reads are observed for lineage i at time point t_k if the lineage is neutral during the time interval (t_{k-1}, t_k), we employ a neutral model that is analogous to the model with selection described above in Sections 3.2.1 and 3.2.2.

As in the model with selection, we start with the belief distribution $P^{\text{belief}}\left(n, s \,|\, \boldsymbol{r}'\right)$, where $\boldsymbol{r}' = (r_0, \ldots, r_{k-1})$ is the observation vector up to and including the previous time point t_{k-1}. Since the neutral model is formulated only in terms of the lineage size n, we first need to obtain the marginal belief distribution $P^{\text{belief}}(n \,|\, \boldsymbol{r}')$. This distribution has no analytical expression. Therefore, to facilitate further calculations, we approximate it with a Gamma distribution with the scale parameter $\tilde{\theta}$ and shape parameter $\tilde{\varkappa}$ chosen to match the marginal mean and variance given by equations (S52), (S53) i.e., we set

$$P^{\text{belief}}(n \,|\, \boldsymbol{r}') = \gamma(n; \tilde{\varkappa}, \tilde{\theta}),$$
$$\text{with } \tilde{\varkappa} = \frac{\varkappa}{(1+\varkappa)e^{b^2} - \varkappa},$$
$$\text{and } \tilde{\theta} = \frac{a e^{\frac{b^2}{2}}}{\tilde{\varkappa}}.$$

We then obtain the prior distribution for the lineage size at the current time point t_k analogously to equation (S44),

$$\tilde{P}^{\text{prior}}(n) = \begin{cases} 1 - \tilde{p}_{\text{surv},L}\,, & \text{if } n = 0, \\ \tilde{p}_{\text{surv},L}\, \gamma\left(n; \tilde{\varkappa}_L, \tilde{\theta}_L\right), & \text{if } n > 0, \end{cases} \tag{S65}$$

where the L-cycle survival probability $\tilde{p}_{\text{surv},L}$ is given by equation (S45) with parameters $\tilde{\theta}_\ell = f_\theta\left(0, \tilde{\theta}_{\ell-1}, \tilde{\varkappa}_{\ell-1}\right)$, $\tilde{\varkappa}_\ell = f_\varkappa\left(\tilde{\theta}_{\ell-1}, \tilde{\varkappa}_{\ell-1}\right)$ for $\ell = 1, 2, \ldots L$ that are

obtained recursively using equations (S42) and (S43) with $\varkappa_0 \equiv \tilde{\varkappa}$, $\theta_0 \equiv \tilde{\theta}$. The marginal likelihood under the neutral model is then given by

$$Z^N\left(\boldsymbol{r}; \bar{s}_k, \epsilon_k\right) = \int_0^\infty \tilde{P}^{\text{prior}}\left(n \,|\, \boldsymbol{r}'\right)\, P^{\text{meas}}(r_k \,|\, n)\, dn. \tag{S66}$$

Numerical estimation of the marginal likelihoods Z^{sel} and Z^{neut}. For every putatively adapted lineage, we estimate the marginal likelihood Z^{sel} given by equation (S47) by first obtaining a random sample (s_j, n_j), $j = 1, 2, \ldots, M$ with $M = 5,000$ from the projected distribution $P^{\text{prior}}(n, s; \bar{s})$ for that lineage, which is given by equation (S44), using a standard random-number generator, keeping only those samples with $n_j > 0$, i.e., conditional on lineage survival. We then calculate

$$\hat{Z}^{\text{sel}}(\boldsymbol{r}; \bar{s}, \epsilon) = \frac{1}{M} \sum_{j=1}^{M} P^{\text{meas}}(r_k \,|\, n_j; \epsilon). \tag{S67}$$

For every putatively neutral lineage, we estimate the marginal likelihood Z^{neut} given by equation (S66) analogously, by first obtaining a random sample n_j, $j = 1, 2, \ldots, M$ from the projected distribution $\tilde{P}^{\text{prior}}(n; \bar{s})$ given by equation (S65), keeping only those samples with $n_j > 0$. We then estimate $\hat{Z}^{\text{neut}}$ as per equation (S67).

3.4 Identification of adapted lineages and estimation of their selection coefficients

As a result of applying BASIL, we obtain a time-varying belief distribution $P_{ik}^{\text{belief}}(n, s) \equiv P_{ik}^{\text{belief}}(n, s|\boldsymbol{r}_{ik})$ for each lineage i at each sampling time point t_k, $k = 1, 2, \ldots$. Our primary interest is in the marginal belief distribution for the selection coefficient s, $P_{ik}^{\text{belief}}(s) \equiv P^{\text{belief}}(s|\boldsymbol{r}_{ik})$ whose mean and variance are μ_{ik} and σ_{ik}^2, estimated using equations (S48), (S49). At early time points, we expect these belief distributions to have high variance because they are insufficiently constrained by the data. As data accumulates, the uncertainty in s should decline, particularly for lineages that actually acquired a single adaptive mutation (Figure S4). However, if we wait long enough, secondary adaptive mutations might appear, which could again increase the uncertainty in our belief of s. Thus, for each lineage i, we find the time point $t_{k'(i)}$ with minimal variance, $k'(i) = \arg\min_k \hat{\sigma}_{ik}^2$. We call lineage i adapted if $\hat{\mu}_{ik'(i)} > \beta\, \hat{\sigma}_{ik'(i)}$ where β is the "confidence factor", a hyper-parameter that controls the precision and recall of our inference. As described in the main text, we empirically determine that the confidence factor $\beta = 3.3$ maximizes the F1-score (the harmonic mean of precision and recall) in our simulated data, and use this value for all our analyses. For each lineage i called as adapted, we estimate its selection coefficient as $\hat{s}_i = \hat{\mu}_{ik'(i)}$, and we can calculate the credible interval for it using the standard deviation $\hat{\sigma}_i = \hat{\sigma}_{ik'(i)}$.

Algorithm 1 BASIL

procedure CONSTRUCT SELECTION MODEL ▷ Sec. 3.2
procedure CONSTRUCT NEUTRAL MODEL ▷ Sec. 3.3
procedure ESTIMATE BELIEF DISTRIBUTIONS
  for all lineage i **do**
    procedure INITIALIZE BELIEF DISTRIBUTION ▷ Sec. **??**
      $P^{\text{belief}}(n, s \,|\, r_{i0}) = N(s; \mu, \sigma)\, \gamma\,(n; \varkappa(r_{i0}), \theta(r_{i0}))$
      with $\mu = 0$, $\sigma = 0.1$, $\varkappa = r_{i0} + 1$, $\theta = (R_0 + N_L)^{-1}$
  for time point $k = 1, 2, \ldots, k_{\text{final}}$ **do**
    procedure ESTIMATE MEAN FITNESS FOR (t_{k-1}, t_k) ▷ Sec. 3.3
      $\bar{s}_k = \arg\max_{\bar{s}} L_k(\bar{s})$ ▷ Eq. (S63)
    for all lineage i **do**
      procedure OBTAIN PRIOR DISTRIBUTION FOR t_k ▷ Sec. 3.2.1
        $P^{\text{belief}}(n, s \,|\, \mathbf{r}_{ik-1}) \to P^{\text{prior}}(n, s \,|\, \mathbf{r}_{ik-1})$ ▷ Eq. (S44)
      procedure UPDATE BELIEF DISTRIBUTION AT t_k ▷ Sec. 3.2.2
        $P^{\text{prior}}(n, s \,|\, \mathbf{r}_{ik-1}) \to P^{\text{belief}}(n, s \,|\, \mathbf{r}_{ik})$ ▷ Eq. (S46)
procedure IDENTIFY ADAPTED LINEAGES ▷ Sec. 3.4
procedure ESTIMATE SELECTION COEFFICIENTS ▷ Sec. 3.4

3.5 Implementation

Figure 3 in the main text and Algorithm 1 show the overall BASIL workflow. We implement BASIL in a Python-based software package. For MCMC sampling, this package uses the C library *stan* and the Python package *pystan 2* as an interface between Python and *stan*. MCMC sampling is implemented using the No-U-Turn Sampler (NUTS) sampler, which efficiently generates proposals based on the posterior distribution [5]. For each belief distribution, we obtain $M = 3500$ samples, with a burn-in of 1,000 steps. For the maximum likelihood estimation of mean fitness, we use the Python package *Noisyopt* for the optimization of noisy functions. Noise arise from the fact that we estimate $\hat{Z}^{\text{sel}}$ and $\hat{Z}^{\text{neut}}$ for every lineage using a Monte Carlo method as described in Section 3.3. The *Noisyopt* package repeatedly evaluates the stochastic target function, then averages over the stochasticity to ensure convergence.

The BASIL analysis of a single BLT experiment with 5×10^5 lineages sampled at 10 time points took us about 50 hours to complete on AMD Ryzen 5 7600X 6-Core Processor, and we found that the runtime increases linearly in the number of lineages and sampling time points. However, since MCMC samples for different lineages are independent at each time point, our algorithm is easily parallelizable, and a multiprocessing capability has been implemented in BASIL. The number of

processors can be set manually depending on the user's environment. In our analyses, we found it convenient to use 12 to 32 processor cores.

BASIL code, example data and installation instructions are available at

https://github.com/HuanyuKuo/BASIL-public

When barcode read count data are provided, the algorithm automatically detects the total number of barcodes and the cycle duration between sampling times. Other parameters can be adjusted in myConstant.py. These include:

- Biological settings: dilution factor, carrying capacity, and the number of randomly chosen reference lineages for mean fitness estimation.
- System settings: number of processors (CPU cores) used for parallel computation.

For adapted lineage calling, we use a confidence factor of $\beta = 3.3$ as the default setting in this work. However, users may adjust β to explore the sensitivity of lineage calls (see Discussion in the main text and Figure S12). Since β only affects the final lineage-calling step, changing its value does not require re-running the time-consuming MCMC estimation.

After completing a run, BASIL generates the following output files:

- Bayesian_global_parameters_XXX: inferred mean fitness trajectory and inferred ϵ trajectory.
- BASIL_Selection_Coefficient_for_called_Adapted_XXX: list of adapted lineages, including barcode indices, estimated selection coefficients (s) with mean and standard deviation, and the calling time.
- posterior_XXX_SModel_S_T1: the parametric belief distributions of all lineages at a particular sampling time point.
- glob_XXX_T1: information of maximizing the likelihood of the mean fitness and ϵ at a particular sampling time point.

We recommend that users first run BASIL on the provided example BLT dataset. This test run helps verify that the installation is working properly and familiarizes users with the workflow before applying the algorithm to their own data.

4 Application of BASIL to simulated data

4.1 Belief distribution and lineage identification

Figure S4 shows how the belief distribution $P_{ik}^{\text{belief}}(n, s)$ changes over time as information about the lineage frequency is accrued. Next, as discussed in the main text, we used simulated data to determine the optimal value of the confidence factor β, setting $\beta = 3.3$. The confidence factor has a simple interpretation. We use a linear classifier to classify lineages into adapted and neutral. Specifically, if we plot the inferred selection coefficient of each lineage $\hat{s}_i$ and the estimated standard deviation $\hat{\sigma}_i$ of the credible interval around it, then $1/\beta$ is the slope of the line that separates adapted from non-adapted lineages (see Figures S5). We find that lineages form two large clusters (with some finer clustering structure visible in Figures S5B) that contain overwhelmingly adapted or non-adapted lineages, respectively. The classification line $\hat{s}_i = \beta\,\hat{\sigma}_i$ with $\beta = 3.3$ separates these two clusters very well (see Table 1 in the main text).

4.2 Performance on simulated data

To analyze the performance of BASIL on lineage identification, we calculate the following standard statistics:

- The number of true positives (TP), i.e., adapted lineages identified as such;
- The number of false positives (FP), i.e., neutral lineages incorrectly identified as adapted;
- The number of true negatives (TN), i.e., neutral lineages identified as such;
- The number of false negatives (FN), i.e., adapted lineages incorrectly identified as neutral;
- Precision, i.e., the fraction of all positives that are true, TP/(TP+FP);
- Recall, or the true positive rate, i.e., the fraction of all adapted lineages that are identified as such, TP/(TP+FN);
- The F1-score, which is the harmonic mean of precision and recall.

All numbers and statistics are reported in Table 1 in the main text. In particular, in the weak selection region, we obtain 31 false positives (precision 98.8%) and 430 false negatives (recall 85.7%), and in the strong selection region, we obtain 7 false positives (precision 99.7%) and 519 false negatives (recall 82.7%). Figure S5 reveals that all false positives, i.e., neutral lineages that are incorrectly identified as

adapted, are very close to the classification line, suggesting that in principle their rate could be further reduced by increasing β (albeit at the expense of increasing the rate of false negatives). In contrast, false negatives, i.e., adapted lineages that are incorrectly classified as neutral, are distributed broadly within the non-adapted cluster, indicating that no linear classifier that is based solely on $\hat{s}_i$ and $\hat{\sigma}_i$ can achieve a substantially higher recall. This observation suggests that false negative lineages may be statistically indistinguishable from true neutral ones.

To investigate this conjecture, we first plotted the trajectories of all lineages stratified by their predicted and actual class labels. Figures S6 and S7 confirm that the trajectories of false negative lineages are visually indistinguishable from those of truly neutral lineages. One possible explanation for why some adapted lineages behave is if they are neutral is that these lineages have such small population sizes that their dynamics are governed largely by genetic drift, that is, they fail to "establish" [6]. An adapted lineage establishes approximately when its size n_i exceeds the inverse of its selection coefficient $1/s_i$, or equivalently, $n_i s_i > 1$ [6]. Thus, lineages with smaller selection coefficients have a smaller chance of successfully establishing. Consistent with this prediction, we find that false negative lineages have significantly smaller selection coefficients than true positive lineages, with $P < 0.001$ in both cases (Welch's t-test; see Figures S8). Next, for each lineage i, we estimated the maximum size it has achieved during the course of the simulation as $\tilde{n}_i = \max_k r_{ik} N_b / R_k$ and plotted the lineage's true selection coefficient s_i against $\tilde{n}_i$. Figure S8 show that for the majority of false negative lineages, we have $\tilde{n}_i s_i < 1$, whereas for the vast majority of true positive lineages, $\tilde{n}_i s_i > 1$, as expected. These observations support our hypothesis that some adapted lineages fail to establish. As a result, their dynamics are highly stochastic and essentially indistinguishable from neutral.

Appendix A Expectation of the noise variance estimator and bias correction

Since the estimate (S4) we calculate its expectation to determine whether it is biased.

$$\mathbb{E}\left(\hat{\sigma}_{pi}\right) = \frac{1}{|X_p|} \sum_{b \in X_p} \left(\mathbb{E}\left(r_{bi}^2\right) - 2\mathbb{E}\left(r_{bi}^2\,\bar{r}_{bi}^2\right) + \mathbb{E}\left(\bar{r}_{bi}^2\right)\right). \tag{S68}$$

To calculate the expectations in the sum of equation (S68), note that

$$\mathbb{E}\left(r_{ib}\,r_{jb}\right) = p_b^2 R_i R_j + \delta_{ij}\,\sigma_{pi}^2.$$

Now, we have

$$\mathbb{E}\left(r_{bi}^2\right) = \left(p_b R_i\right)^2 + \sigma_{pi}^2, \tag{S69}$$

$$\mathbb{E}\left(r_{bi}^2\,\bar{r}_{bi}^2\right) = \frac{R_i}{n_{\text{rep}}} \sum_{j=1}^{n_{\text{rep}}} \frac{\mathbb{E}\left(r_{bi}\,r_{bj}\right)}{R_j} = \left(p_b R_i\right)^2 + \frac{\sigma_{pi}^2}{n_{\text{rep}}}, \tag{S70}$$

$$\mathbb{E}\left(\bar{r}_{bi}^2\right) = \left(\frac{R_i}{n_{\text{rep}}}\right)^2 \sum_{j,k=1}^{n_{\text{rep}}} \frac{\mathbb{E}\left(r_{bj}\,r_{bk}\right)}{R_j R_k} = \left(p_b R_i\right)^2 + \left(\frac{R_i}{n_{\text{rep}}}\right)^2 \sum_{j=1}^{n_{\text{rep}}} \frac{\sigma_{pj}^2}{R_j^2}. \tag{S71}$$

Substituting equations (S69)–(S71) into equation (S68), we obtain

$$\mathbb{E}\left(\hat{\sigma}_{pi}\right) = \sigma_{pi}^2 \left(1 - \frac{1}{n_{\text{rep}}}\right)^2 + \left(\frac{R_i}{n_{\text{rep}}}\right)^2 \sum_{j \neq i} \frac{\sigma_{pj}^2}{R_j^2}. \tag{S72}$$

Thus, the estimate (S4) is a biased estimate of σ_{pi}^2. Note that if all replicates have the same coverage, $R_i = R$, and $\sigma_{pi}^2 = \sigma_p^2$, equation (S72) simplifies to

$$\mathbb{E}\left(\hat{\sigma}_{pi}\right) = \sigma_p^2 \left[\left(1 - \frac{1}{n_{\text{rep}}}\right)^2 + \frac{n_{\text{rep}} - 1}{n_{\text{rep}}^2}\right] = \sigma_p^2 \left(1 - \frac{1}{n_{\text{rep}}}\right),$$

and correcting the bias is easy. To correct the bias in the general case, we rewrite equation (S72) in the matrix form as

$$\mathbb{E}\left(\boldsymbol{V}_p\right) = A\,\boldsymbol{\sigma}_p^2,$$

where $\boldsymbol{V}_p = \left(\hat{\sigma}_{p1}, \hat{\sigma}_{p2}, \ldots, \hat{\sigma}_{pn_{\text{rep}}}\right)^T$ and $\boldsymbol{\sigma}_p^2 = \left(\sigma_{p1}^2, \sigma_{p2}^2, \ldots, \sigma_{pn_{\text{rep}}}^2\right)^T$ and

$$A = \begin{pmatrix} \left(1 - \frac{1}{n_{\text{rep}}}\right)^2 & \frac{1}{n_{\text{rep}}^2}\frac{R_1^2}{R_2^2} & \cdots & \frac{1}{n_{\text{rep}}^2}\frac{R_1^2}{R_{n_{\text{rep}}}^2} \\ \frac{1}{n_{\text{rep}}^2}\frac{R_2^2}{R_1^2} & \left(1 - \frac{1}{n_{\text{rep}}}\right)^2 & \cdots & \frac{1}{n_{\text{rep}}^2}\frac{R_2^2}{R_{n_{\text{rep}}}^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n_{\text{rep}}^2}\frac{R_{n_{\text{rep}}}^2}{R_1^2} & \frac{1}{n_{\text{rep}}^2}\frac{R_{n_{\text{rep}}}^2}{R_2^2} & \cdots & \left(1 - \frac{1}{n_{\text{rep}}}\right)^2 \end{pmatrix},$$

and obtain the bias-corrected estimate

$$\boldsymbol{V}'_p = A^{-1}\boldsymbol{V}_p. \tag{S73}$$

Appendix B Expected frequency given the number of reads

Here, we derive the expressions (S22)–(S28) for the conditional expected frequency of a lineage $\mathbb{E}\,(x|r)$, given the number of reads r.

Example 1. Poisson measurement noise and a uniform frequency distribution. We first consider the case when the measurement process is modeled by the Poisson distribution (S21) and the frequency distribution is uniform. Then, the conditional distribution for the frequency x is given by

$$P(x|r)\,dx = \frac{P^{\text{meas}}(r|x)\,P^{\text{prior}}(x)}{P(r)}\,dx = C\,x^r\,e^{-xR}\,dx, \quad \text{with } x \in [0,1], \tag{S74}$$

where C is the normalization constant. As long as $r/R \ll 1$, this distribution is well approximated by the gamma distribution with the shape parameter $k = r+1$ and scale parameter $\theta = 1/R$. Thus, $C = R^{r+1}/r!$ whose mean is $k\theta = (r+1)/R$ confirming (S22).

Example 2. Poisson measurement noise and an exponential frequency distribution. If the frequency distribution is exponential with mean $1/N_L$ (see equation (S23)), then we have

$$P(x|r)\,dx = C\,x^r\,e^{-x(R+N_L)}\,dx, \tag{S75}$$

which is again is very well approximated by the gamma distribution with the shape parameter $k = r+1$ and scale parameter $\theta = 1/(R+N_L)$, as long as $r \ll R$. Expression (S24) follows immediately.

Example 3. Measurement noise with an increasing variance to mean ratio and a Gamma frequency distribution. We next consider the case when the frequency distribution is Gamma with shape parameter α and scale parameter $1/\tilde{N}$ (see equation (S26)) and the read count is a negative binomial random variable with mean xR and the variance to the mean ratio $1+\epsilon xR$ (see (S25)). When $xR \ll 1/\epsilon$, the negative binomial distribution (S25) converges to the Poisson distribution with mean xR, which implies that the conditional distribution for x is a Gamma

distribution with shape parameter $r+\alpha$ and scale parameter $(R+\tilde{N})^{-1}$. More generally, the conditional distribution for x given r is given by

$$P(x|r)dx = \frac{P^{\text{meas}}(r|x)\,P^{\text{prior}}(x)}{P(r)}dx = C\,x^{r+\alpha-1}(1+\epsilon Rx)^{-(1/\epsilon+r)}\,e^{-\tilde{N}x}\,dx. \quad \text{(S76)}$$

It is possible to use Laplace's method to obtain approximate expressions for the normalization constant C and the moments of this distribution. Specifically, whenever $\epsilon \ll 1$, we obtain

$$C = \frac{\tilde{N}+R}{\sqrt{8\pi(r+\alpha-1)}}\,\frac{(1+\epsilon R\bar{x}_0)^{\frac{1}{\epsilon}+r}}{\bar{x}_0^{\alpha+r+1}}\,e^{\tilde{N}\bar{x}_0}$$

$$\mathbb{E}\left(x^i|r\right) = (\bar{x}_i)^i\left(\frac{r+\alpha+i-1}{r+\alpha-1}\right)^{r+\alpha-\frac{1}{2}}\left(\frac{1+\epsilon R\bar{x}_0}{1+\epsilon R\bar{x}_i}\right)^{\frac{1}{\epsilon}+r}e^{-\frac{\tilde{N}i}{\tilde{N}+R}},$$

for $i, r = 1, 2, \ldots$, where $\bar{x}_i = \frac{r+\alpha+i-1}{\tilde{N}+R}$.

Example 4. Measurement noise with a constant variance to mean ratio and an exponential frequency distribution. Finally, we consider the case when the measurement process is over-dispersed but with a constant variance to the mean ratio 2κ (see equation (S25)) and the frequency distribution is exponential (see equation (S23)). To facilitate analytical tractability, instead of tracking frequency x of the lineage in the population, we will track its lineage size n. We assume that n take values $1, 2, \ldots$ and that its prior distribution is geometric with parameter q. The expected lineage size is N/N_L, which implies that $q = N_L/N$. Thus, for the conditional distribution for n, given r is

$$P(n|r) = \frac{P^{\text{meas}}(r|n)\,P^{\text{prior}}(n)}{P(r)} = \frac{1}{P(r)}\,\frac{\Gamma(r+k)}{\Gamma(r+1)\Gamma(k)}(1-p)^r p^k\,(1-q)^{n-1}q, \quad \text{(S77)}$$

where $p = 1/2\kappa$, $k = \alpha n$, $q = N_L/N$ and $\alpha = R/N(2\kappa-1)^{-1}$. Expression (S77) suggests that the quantity $\tilde{k} = \alpha n - 1$ is a negative binomial random variable with parameters

$$\tilde{r} = r+1, \quad \text{(S78)}$$
$$\tilde{p} = 1 - p\,(1-q)^{1/\alpha},$$

that is,

$$P(\tilde{k}|r) = C\frac{\Gamma\left(\tilde{k}+\tilde{r}\right)}{\Gamma\left(\tilde{k}+1\right)}\,(1-\tilde{p})^{\tilde{k}}, \quad \text{(S79)}$$

where $C = \tilde{p}^{\tilde{r}}/\Gamma(\tilde{r})$ is the normalization constant. Thus, we have

$$\mathbb{E}\left(\tilde{k}\middle|r\right) = \frac{\tilde{r}(1-\tilde{p})}{\tilde{p}} = \frac{p\,(r+1)\,(1-q)^{1/\alpha}}{1-p\,(1-q)^{1/\alpha}}.$$

Since $\tilde{k} = x\frac{R}{2\kappa-1} - 1$, we have

$$\mathbb{E}\left(x|r\right) = \frac{2\kappa-1}{R}\frac{2\kappa+r\,(1-q)^{1/\alpha}}{2\kappa-(1-q)^{1/\alpha}} \approx \frac{2\kappa+r}{R},$$

where the last approximation holds because $q = N_L/N \sim 0.01 \ll 1$.

It will also be useful to calculate the expected number of cells in a lineage whose measured number of reads is r,

$$\mathbb{E}\left(n|r\right) = \frac{\mathbb{E}\left(\tilde{k}|r\right)+1}{\alpha} \approx \frac{N}{R}\left(2\kappa+r\right), \tag{S80}$$

and its variance

$$\begin{aligned}\mathrm{Var}\left(n|r\right) &= \frac{1}{\alpha^2}\,\mathrm{Var}\left(\tilde{k}|r\right) = \frac{1}{\alpha^2}\,\frac{\tilde{r}(1-\tilde{p})}{\tilde{p}^2} \\ &= \frac{1}{\alpha^2}\,\frac{(r+1)p\,(1-q)^{1/\alpha}}{(1-p\,(1-q)^{1/\alpha})^2} \approx \frac{N^2}{R^2}\,2\kappa\,(r+1)\,.\end{aligned} \tag{S81}$$

Appendix C Expectation of the logarithm of a gamma-distributed random variable

Consider a random variable X that is Gamma-distributed with scale parameter θ and shape parameter $\varkappa$, such that its probability density is

$$p(x) = \frac{1}{\Gamma(\varkappa)\,\theta^{\varkappa}}\,x^{\varkappa-1}\,e^{-\frac{x}{\theta}}.$$

Therefore, we have

$$\int_0^\infty x^{\varkappa-1}\,e^{-\frac{x}{\theta}}\,dx = \Gamma(\varkappa)\,\theta^{\varkappa}. \tag{S82}$$

Differentiating both sides of equation (S82) with respect to $\varkappa$, we obtain

$$\int_0^\infty \ln x\,x^{\varkappa-1}\,e^{-\frac{x}{\theta}}\,dx = \Gamma(\varkappa)\,\theta^{\varkappa}\left(\frac{\Gamma'(\varkappa)}{\Gamma(\varkappa)} + \ln\theta\right),$$

which can be rewritten as

$$\mathbb{E}\left(\ln X\right) = \psi(\varkappa) + \ln\theta,$$

where $\psi(\varkappa) = \frac{\Gamma'(\varkappa)}{\Gamma(\varkappa)}$ is the digamma function.

5 Supplementary Figures

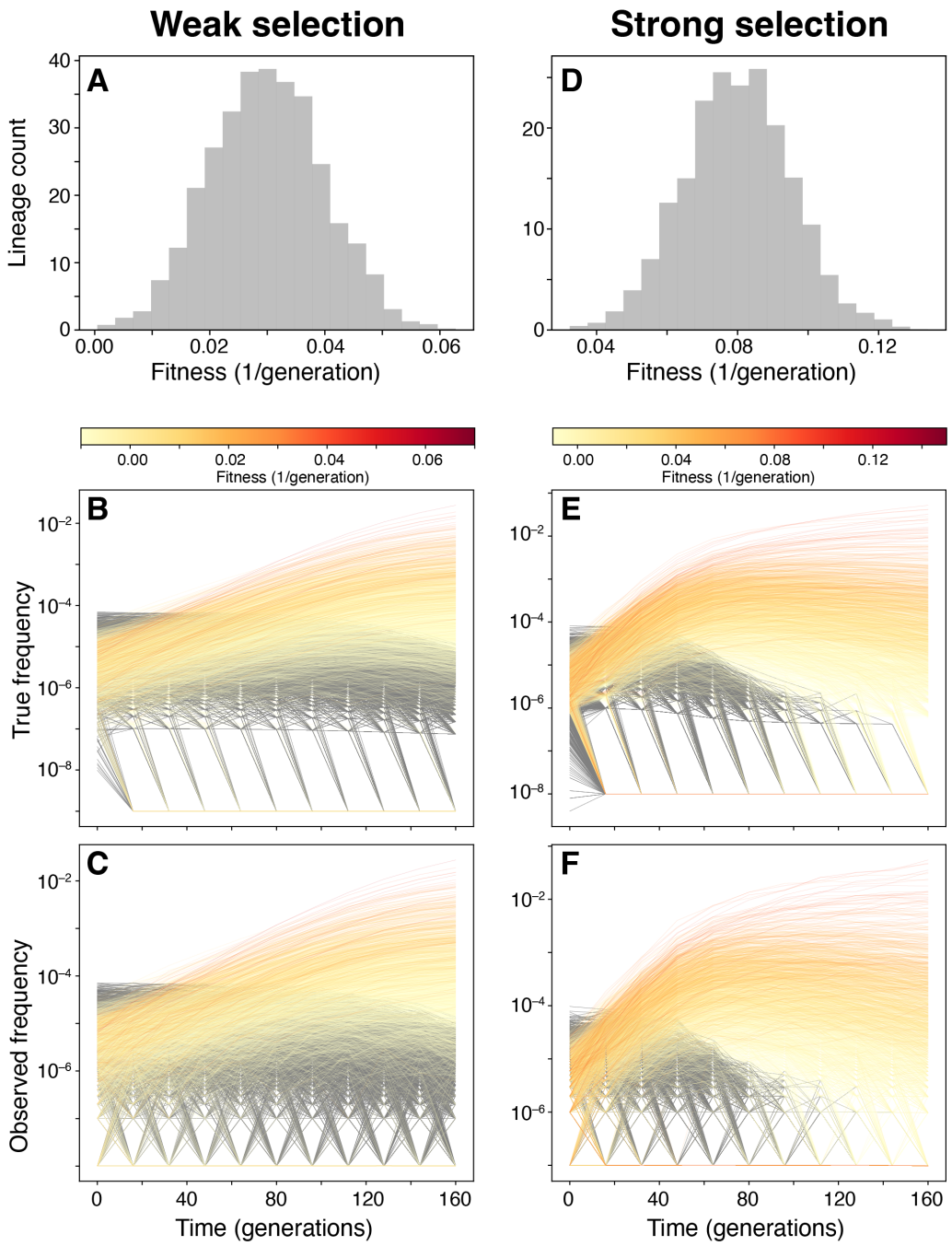

Figure S1. Simulated data. A–C. Weak selection regime. **D–F.** Strong selection regime. Distributions of lineage fitness (A,D). True lineage frequency trajectories (B,E). Observed barcode frequency trajectories (C,F). Adapted lineages are colored by their fitness, neutral lineages are gray.

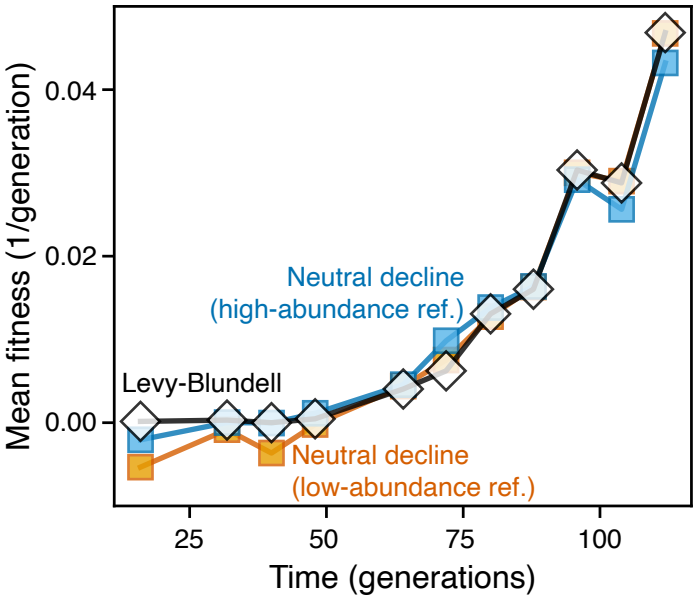

Figure S2. Correspondence between the full Levy-Blundell inference method and the simplified neutral decline method. Mean-fitness trajectories in Replicate 1 in the Levy 2015 dataset as reported in Ref. [3] (black line and white diamonds), and inferred by our the neutral decline method using either high-abundance lineages (blue line and squares) or low-abundance lineages (orange line and squares) as reference.

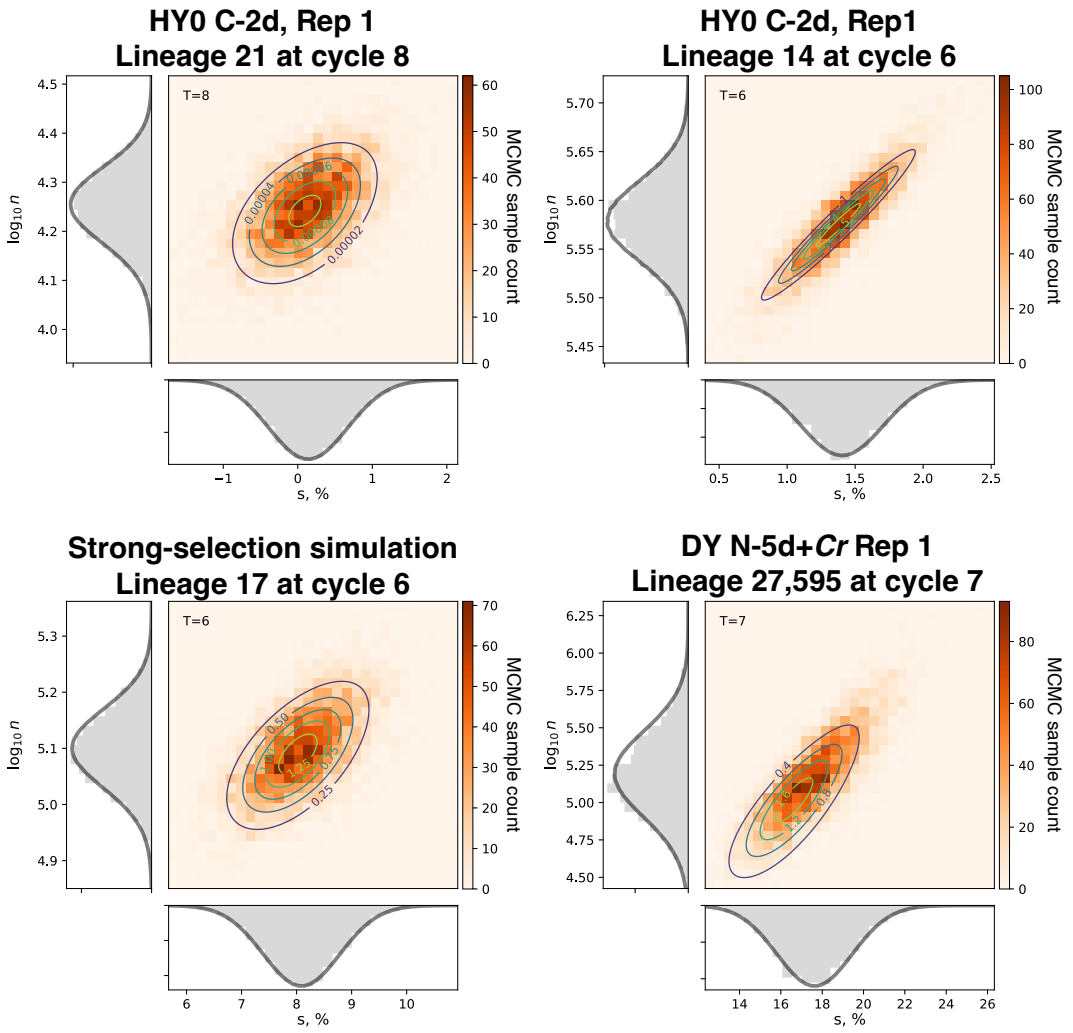

Figure S3. Estimation of the belief distribution $P^{\mathrm{b}}(n, s)$. Each panel corresponds to a single lineage in a particular dataset at a certain time point, as indicated in the panel title. The heatmap in each panel shows the number of MCMC samples in each $(s, \log_{10} n)$ bin. The contour lines correspond to the analytical function $P^{\mathrm{b}}(n, s)$ given by equations (S31)–(S33) with parameters fitted to the MCMC sample. Histograms show the marginal distributions for s (bottom) and $\log_{10} n$ (left) and the corresponding parametric curves. See Section 3.2.2 for details.

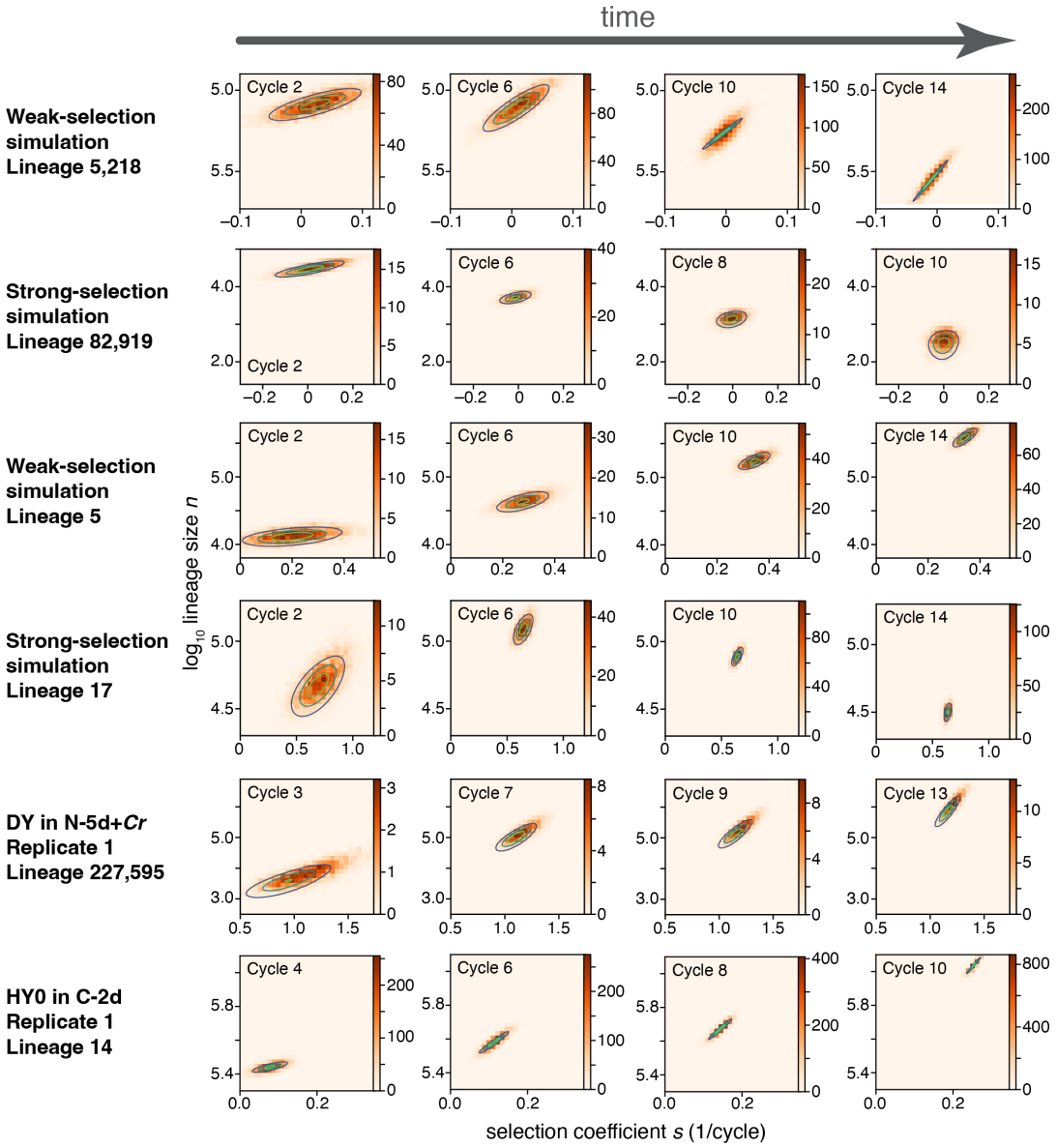

Figure S4. Changes in the belief distribution as data is accumulated. Each row corresponds to a single lineage in a particular dataset, as indicated on the left. Note that the selection coefficient here is shown on the per cycle basis. Notations are as in Figure S3.

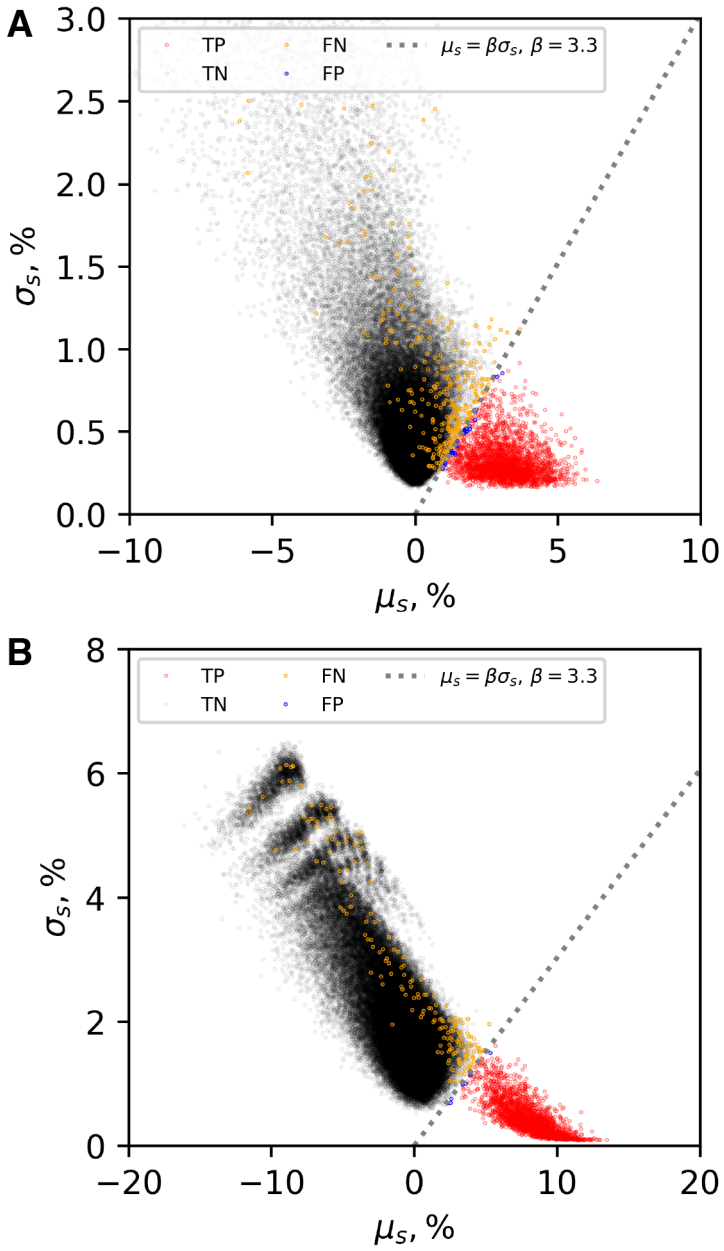

Figure S5. Classification of lineages in simulated data. A. Mean (x axis) versus the standard deviation (y axis) of the estimated marginal belief distribution for the lineage selection coefficient for simulated data in the weak selection regime. **B.** Same for the strong-selection regime. Each point represents a lineage. Colors represent different lineage classes: red = true positives, black = true negatives, orange = false negatives, blue = false positives. Dashed line is the classification line with $\beta = 3.3$.

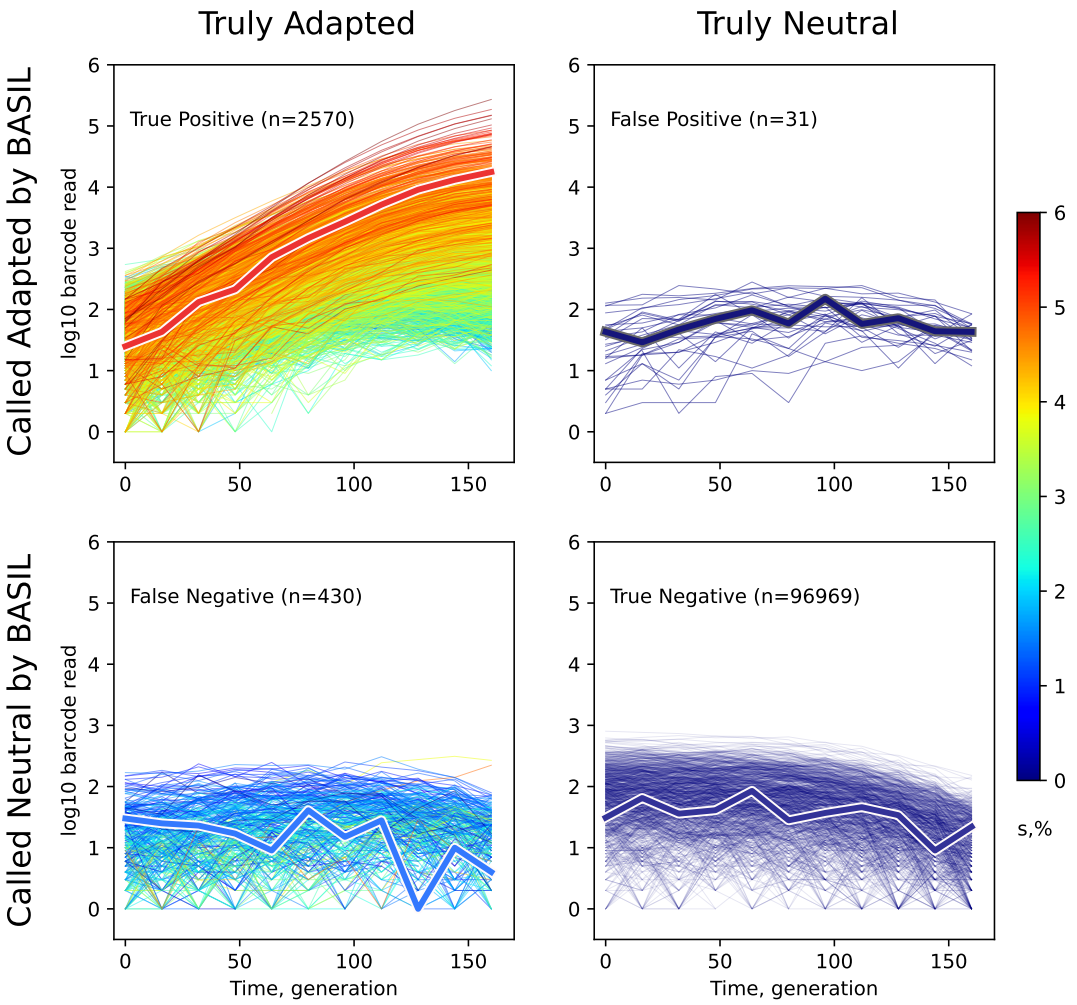

Figure S6. Lineage trajectories in the weak-selection simulation stratified by class. Data is the same as in Figure S1B. Only 970 random true negative lineages are shown to improve visualization. Trajectories are colored by lineage fitness.

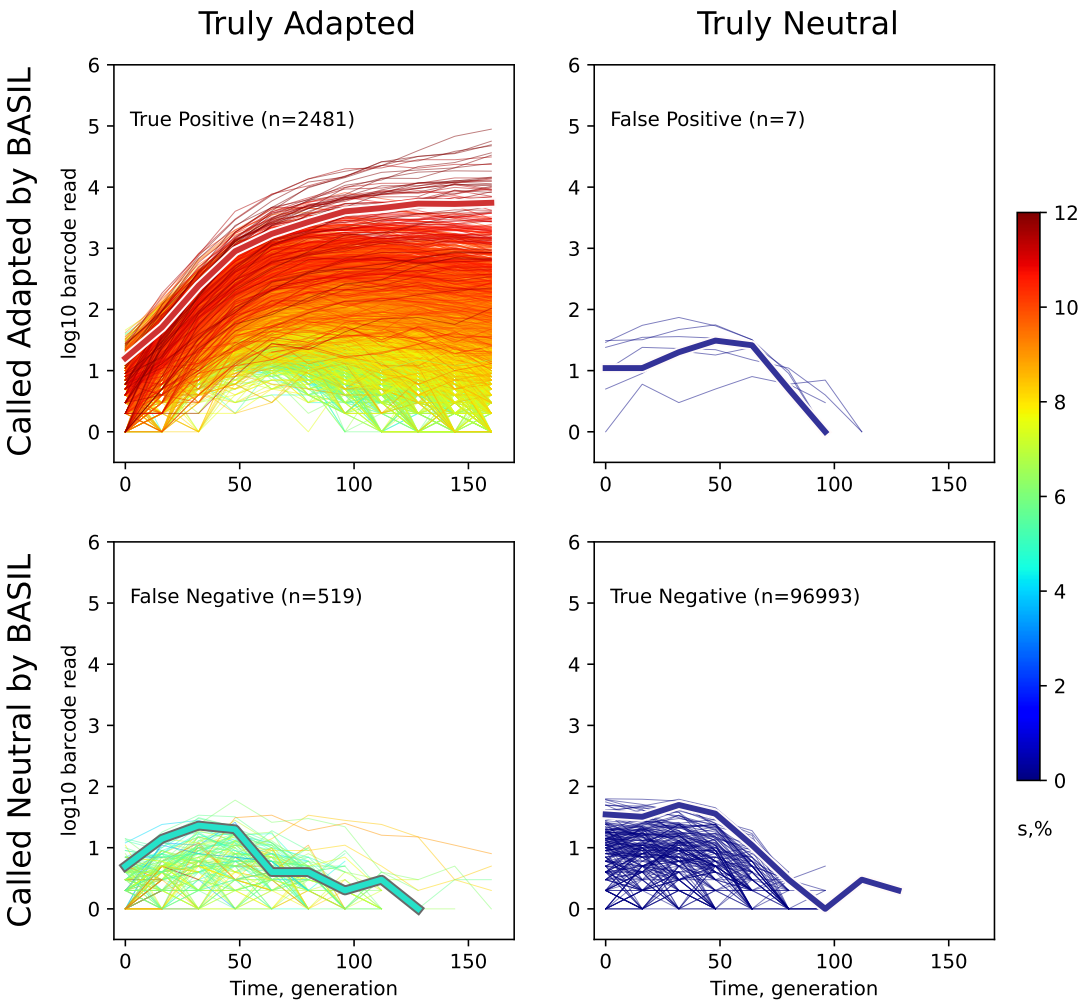

Figure S7. Lineage trajectories in the strong-selection simulation stratified by class. Data is the same as in Figure S1E. Only 243 random true negative lineages are shown to improve visualization. Trajectories are colored by lineage fitness.

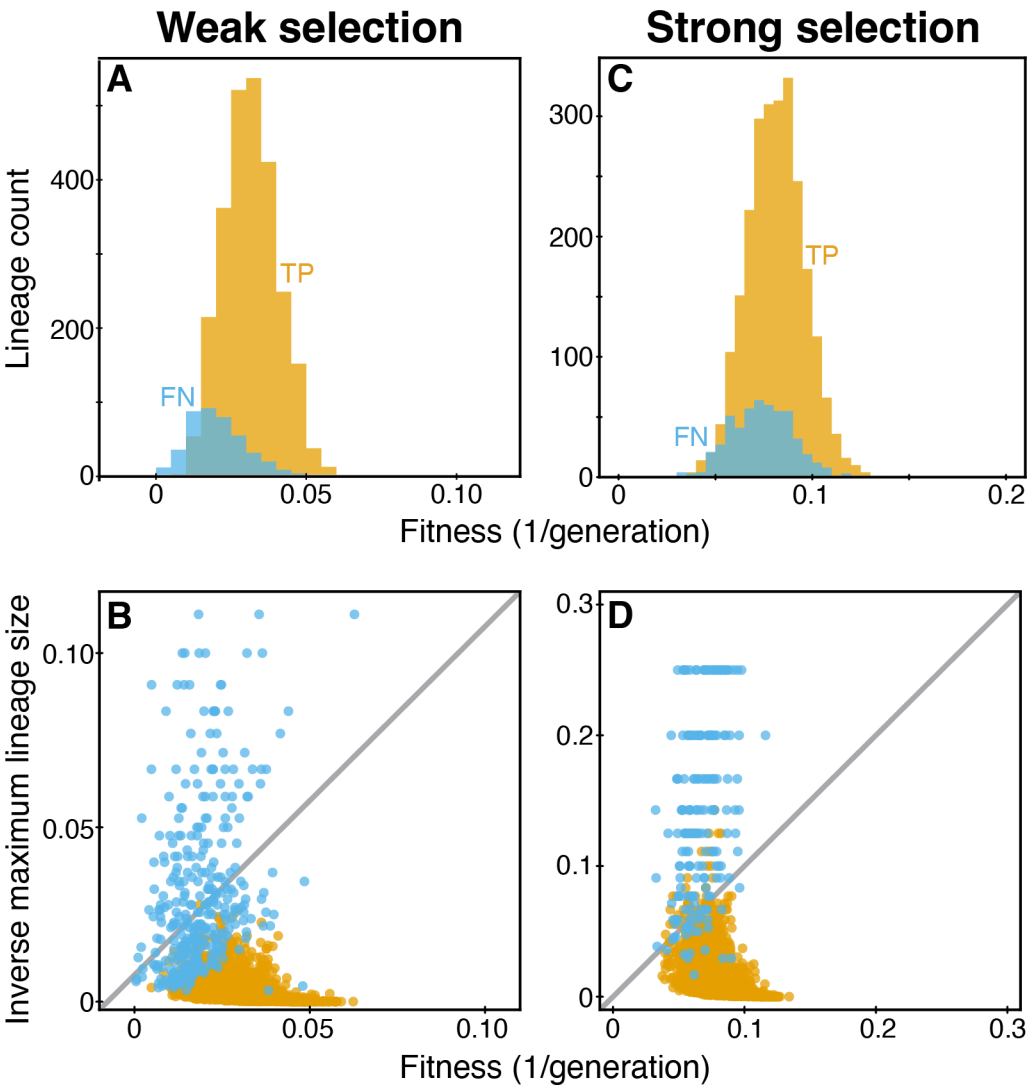

Figure S8. True positive and false negative lineages have distinct features in simulated data. A. Distribution of selection coefficients of true positives (yellow) and false negative (blue) lineages in the weak-selection simulation. **B.** The relationship between lineage fitness s_i and its maximum size $\tilde{n}_i$ in the simulation (see Section 4.2 for details). Lineages above the diagonal (gray line) are those with $\tilde{n}_i s_i < 1$, and lineages below the diagonal are those with $\tilde{n}_i s_i > 1$. **C.** Same as panel A but for the strong-selection simulation. **D.** Same as panel B but for the strong-selection simulation.

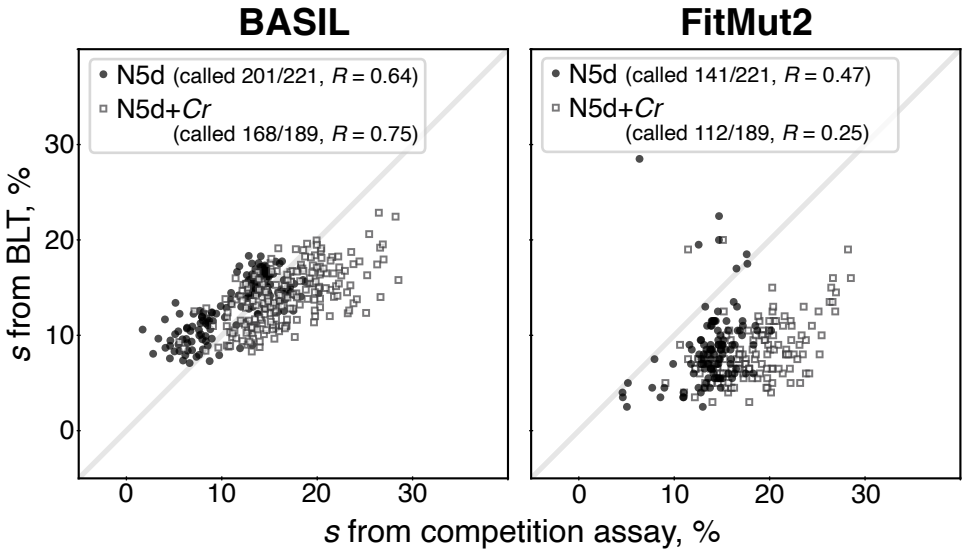

Figure S9. Comparison of BASIL and FitMut2 performance on Venkataram 2023 data. Lineage fitness inferred from BLT experiments by either BASIL or FitMut2 is plotted against the fitness of respective isolated clones measured in competition assays by Venkataram et al [7]. Filled circles represent lineages/clones measured in the N5d condition, which Venkataram et al refer to as "Alone". Empty squares represent lineages/clones measured in the N5d+*Cr* condition, which Venkataram et al refer to as "Community". The number of lineages called as adapted in the respective BLT experiment as well as Pearson correlation coefficient R are shown in parenthesis.

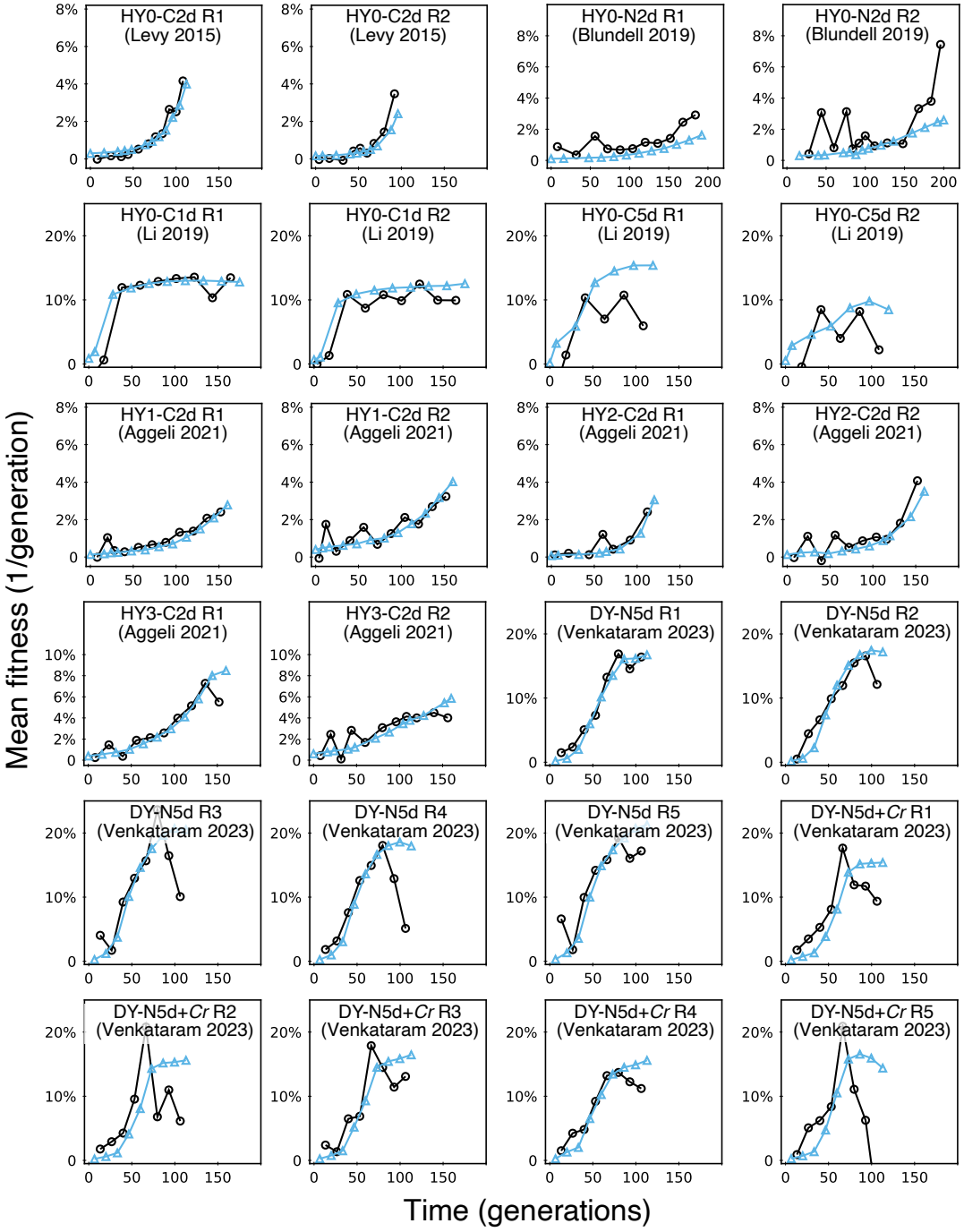

Figure S10. Mean fitness trajectories in published BLT datasets. Each panel represents a replicate of a BLT experiment, as indicated. Black lines show trajectories inferred by BASIL; blue lines show trajectories calculated from lineages identified as adapted.

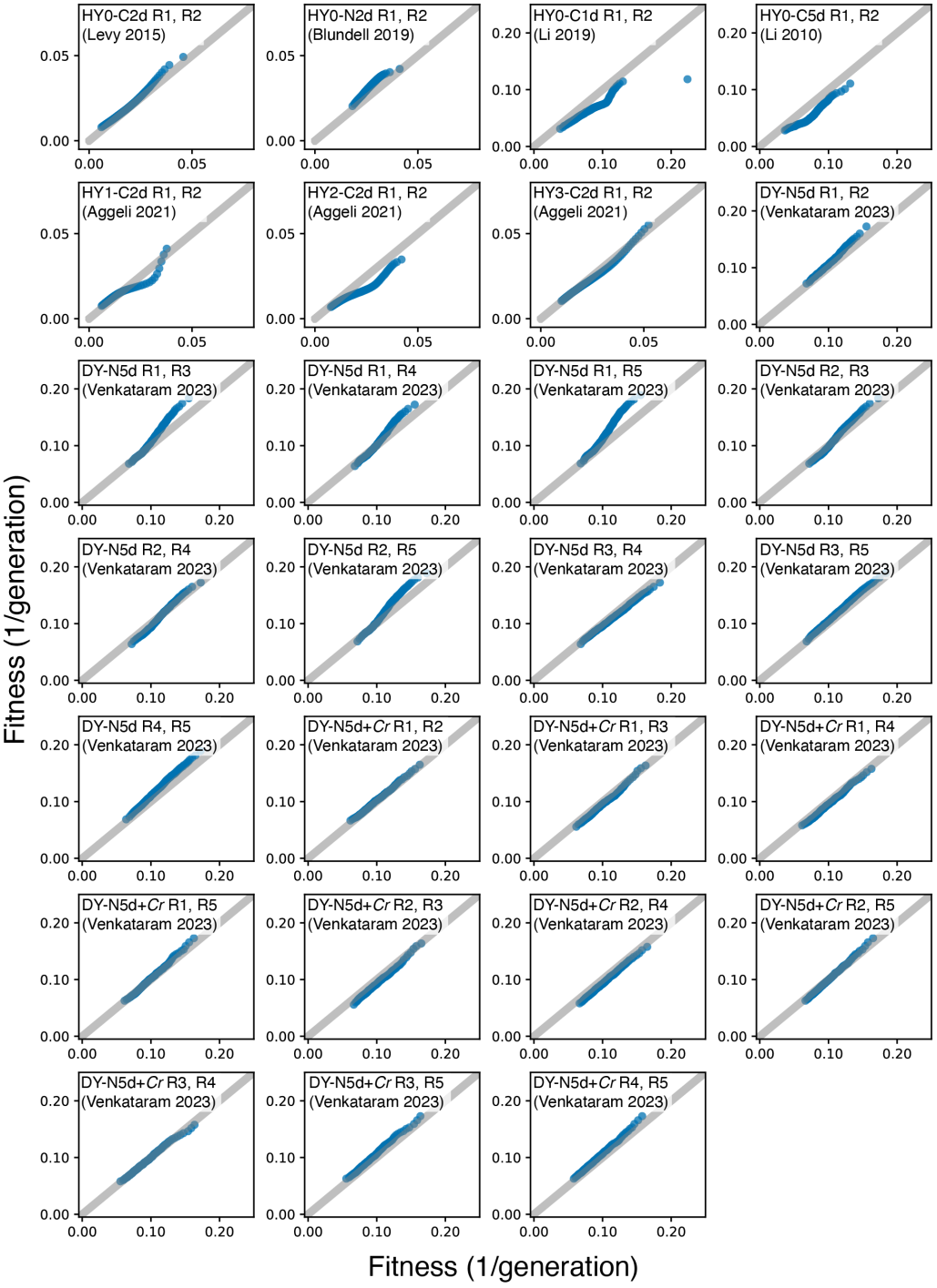

Figure S11. Quantile-quantile plots for the measured distributions of fitness effects (mDFEs) inferred across replicates. Each panel represents a pair of replicates from the same BLT study, as indicated. 100 quantiles are shown in each panel.

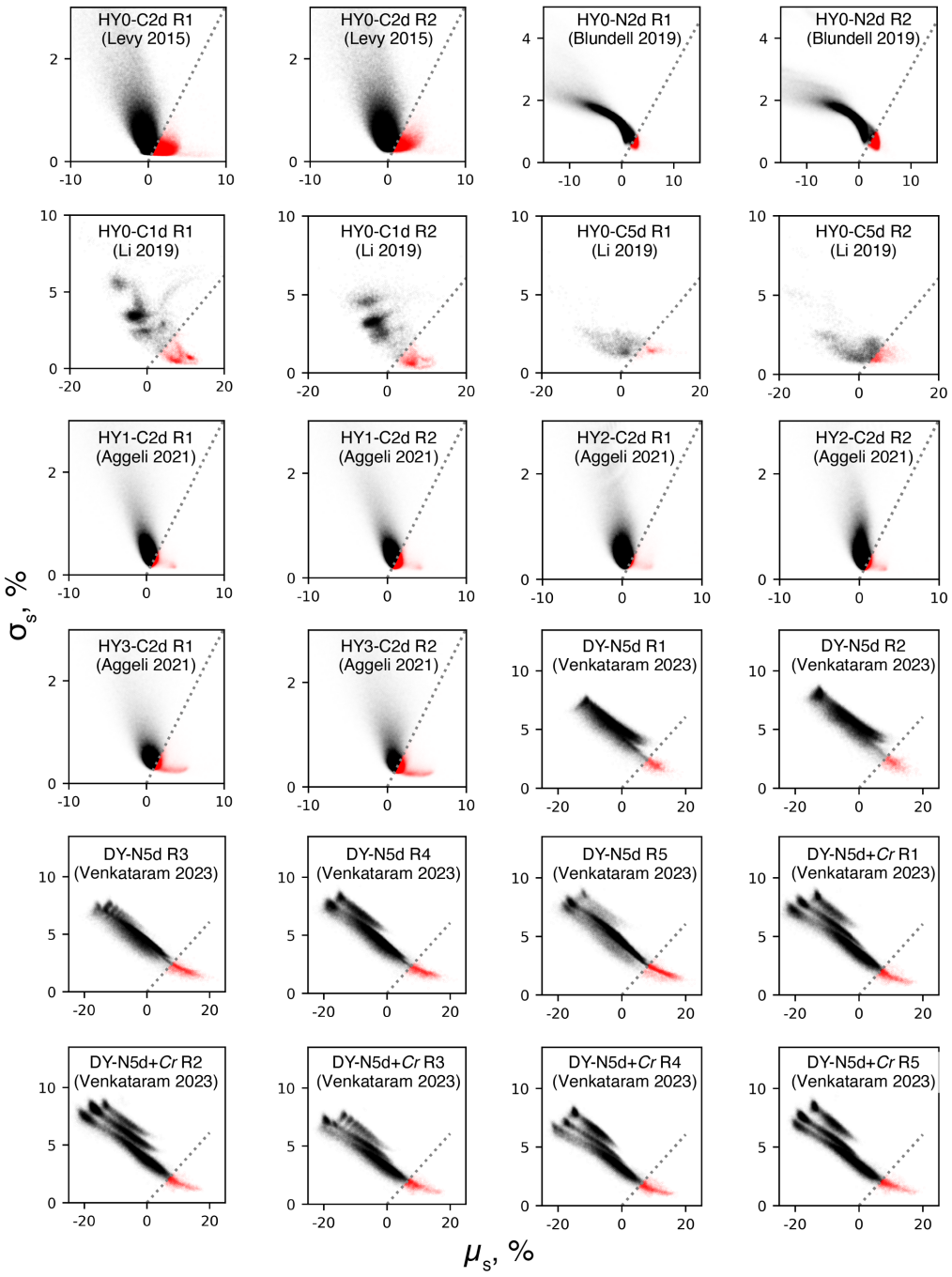

Figure S12. Classification of lineages in real BLT data. Each panel represents a replicate of a BLT experiment, as indicated. Notations are as in Figure S5.

6 Supplementary Tables

	Linear ($\epsilon = 0$)		Quadratic ($\epsilon \neq 0$)	
	Poisson	General	1-param	General
	($a = 1$)	–	($a = 1$)	–
Fitted $\hat{a}$	–	4.34	–	1.97
Fitted $\hat{\epsilon}$	–	–	0.0024	0.0016
RSS[1]	173.6	85.2	55.1	48.3
R^2	0.77	0.89	0.93	0.94

Table S1. Model fitting for the relationship between read-count mean and variance. [1]Residual sum of squares.

Null model			Alternative model				
Name	p_{null}	RSS_{null}	Name	p_{alt}	RSS_{alt}	F	P-value
Poisson	0	173.6	Lin. Gen.	1	85.2	41.5	1.13×10^{-7}
Poisson	0	173.6	Quad. 1-param.	1	55.1	50.7	1.63×10^{-11}
Poisson	0	173.6	Quad. Gen.	2	48.3	86.0	1.44×10^{-11}
Lin. Gen.	1	85.2	Quad. Gen.	2	48.3	29.9	2.87×10^{-6}
Quad. 1-param.	1	55.1	Quad. Gen.	2	48.3	5.56	0.0235

Table S2. Goodness of fit comparison between models fitting the relationship between read-count mean and variance. In all cases $n = 41$.

Parameter	Selection	
	Weak	Strong
$\langle s \rangle$	0.03	0.08
σ_s	0.01	0.015
$\langle N \rangle$	2.56×10^9	2.56×10^8
$\langle N_b \rangle$	10^7	10^6
D	256	256
R	10^7	10^6
σ_r^2	$6 \langle r_i \rangle$	$\langle r_i \rangle + 0.01 \langle r_i \rangle^2$

Table S3. Simulation parameters. For the weak selection regime, we chose the relationship between mean read count $\langle r_i \rangle$ and variance in read count σ_r^2 to be linear in order to make our simulations as similar to the work by Levy et al [3] as possible. For the strong selection regime, we chose the more realistic quadratic relationship, based on our results.

Table S4. Summary of BLT methods

	Levy-Blundell (LB method)	**FitMut2 (modified LB)**	**BASIL (this work)**
Evolution model	Branching process of population size from a single cell	same as the LB method	Exponential growth and dilution in batch culture
Barcode measuring model	Noise model with a linear meanvariance. parameter: κ	same as the LB method	Negative binomial with a quadratic meanvariance. parameter: ϵ
Probability of n trajectory	Not applied	Not applied	The (Bayesian) belief distribution Sec.3.2
Probability of r trajectory	Noise model description Eq. (S8)	Noise model description with both adapted and neutral models Eq. (S12)	Marginal probability of read count Eq. (S47)
Mean-fitness estimation, $\bar{s}$	Maximizing the likelihood function from the reference lineages	Direct calculation from the adapted lineages	Maximizing the likelihood function from the reference lineages
Choice of reference lineage	All small lineages with $20 < r < 40$, assuming they are putatively neutral	Recursively select all putatively adapted lineages	Randomly select 3000 lineages, then assign them to putatively neutral or adapted group
Neutral decline assumption	Yes	No	No
Lineage identification	Ratio of posterior probabilities of adapted and neutral models Eq. (S11)	Bayesian probability of model choice Eq. (S13)	Uncertainty of s in belief distribution Sec.3.4
Fitness effect inference, s	Maximizing the likelihood function for r trajectory Eq. (S10)	Maximizing the likelihood function for r trajectory	The mean of s in belief distribution Sec.3.4

Notation	Description	Eqns
k	sampling time index, $k \in [0, k_{\mathrm{f}}]$	
t_k	time (in generation) of the kth sample	
R_k	total read depth at t_k	
D	dilution ratio, $D > 1$	
N	population size before dilution	
N_b	population size after dilution, $N_b = N/D$	
ϵ	measurement error parameter	(2), (**??**)
s_i	selection coefficient (fitness) of the ith lineage	
$\bar{s}_k$	mean fitness of the population in time interval (t_{k-1}, t_k) for all $k \geq 1$	
n_{ik}	size of ith lineage at t_k before dilution	
r_{ik}	barcode read count of ith lineage at t_k	
$\boldsymbol{r}_{ik}$	observation vector for lineage i prior to and including t_k, $\boldsymbol{r}_{ik} = (r_{i0}, \ldots, r_{ik})$	
$P^{\mathrm{belief}}(n, s\|\boldsymbol{r})$	belief probability that a lineage with the observation vector $\boldsymbol{r}$ has size n and fitness s	(3), (S46)
$P^{\mathrm{belief}}_{ik}(n, s)$	shorthand for $P^{\mathrm{belief}}(n, s\|\boldsymbol{r}_{ik})$	
$P^{\mathrm{meas}}(r\|n)$	probability of observing r reads for a lineage with size n. P^{meas} depends on the total read depth R and the measurement error parameter ϵ	(S30)
$P^{\mathrm{prior}}(n, s\|\boldsymbol{r}')$	prior probability that a lineage with the observation vector $\boldsymbol{r}'$ has size n and selection coefficient s at the *next* sampling time point. P^{prior} depends on the current mean fitness $\bar{s}$	(S44)
$Z^{\mathrm{neut}}(r_{ik})$	marginal likelihood of r_{ik} in the neutral model. Z^{neut} depends on the current mean fitness $\bar{s}$ and measurement error parameter ϵ	(S66)
$Z^{\mathrm{sel}}(r_{ik})$	marginal likelihood of r_{ik} in the selection model. Z^{sel} depends on the current mean fitness $\bar{s}$ and measurement error parameter ϵ	(S47)
$L_k(\bar{s}, \epsilon)$	log-likelihood function for the mean fitness $\bar{s}$ and measurement error parameter ϵ	(S63)

Table S5. Notations used in BASIL.

Strain	Cond.	Ref.	Mean	Median	IQR	IDR
HY0	C-2d	[3]	0.12*	0.05*	0.02 (0.060)	0.11*
HY0	N-2d	[3]	0.48*	0.51*	0.15*	0.25*
HY0	C-1d	[8]	1.85*	1.74*	1.31*	0.72*
HY0	C-5d	[8]	2.01*	2.35*	0.31 (0.081)	0.15 (0.48)
HY1	C-2d	[9]	0.01*	0.11*	0.13*	0.79*
HY2	C-2d	[9]	0.41*	0.30*	0.60*	0.67*
HY3	C-2d	[9]	0.05*	0.05*	0.22*	0.15*
DY	N-5d	[7]	0.74*	0.66*	0.98*	1.49*
DY	N-5d+Cr	[7]	0.43*	0.43*	0.16 (0.070)	0.34*

Table S6. Differences of mDFE statistics across replicates. For each mDFE statistic, we report the absolute value of the pairwise difference in the statistic value, averaged across all pairs of replicates (see text for details). * indicates P-value < 0.01 (permutation test); P-values ≥ 0.01 are shown in parentheses.

		BASIL	**FitMut2**
Full data	# of adapt. lin.	60920	15386
Reduced coverage	# of adapt. lin.	49330	1131
	Change in calling rate	-19.0%	-98.1%
	Overlap	48738	1130
	Relative overlap	98.8%	99.9%
	s PCC[1]	95.8%	82.0%
	Error[2] in s	0.20%	0.83%
Reduced sampling	# of adapt. lin.	48206	55248
	Change in calling rate	-20.9%	$+259.1\%$
	Overlap	48177	15381
	Relative overlap	99.9%	$100,0\%$
	s PCC[1]	97.9%	62.0%
	Error[2] in s	0.18%	1.29%

Table S7. Effects of down-sampling. The "Full" dataset is Replicate 1 of the HY0 strain in C-1d environment [3]. See Materials and Methods in the main text for details. [1]Pearson correlation coefficient. [2]Averaged absolute difference of s inferred in Full and Reduced data.

References

[1] Venkataram S, Kinsler G. BarcodeCounter2; 2021. https://github.com/sandeepvenkataram/BarcodeCounter2.

[2] Johnson MS, Venkataram S, Kryazhimskiy S. Best practices in designing, sequencing, and identifying random DNA barcodes. J Mol Evol. 2023;91(3):263–280.

[3] Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature. 2015;519(7542):181–186.

[4] Li F, Mahadevan A, Sherlock G. An improved algorithm for inferring mutational parameters from bar-seq evolution experiments. BMC Genomics. 2023;24(1):246.

[5] Hoffman MD, Gelman A, et al. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J Mach Learn Res. 2014;15(1):1593–1623.

[6] Desai MM, Fisher DS. Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics. 2007;176(3):1759–1798.

[7] Venkataram S, Kuo HY, Hom EF, Kryazhimskiy S. Mutualism-enhancing mutations dominate early adaptation in a two-species microbial community. Nat Ecol Evol. 2023;7(1):143–154.

[8] Li Y, Petrov DA, Sherlock G. Single nucleotide mapping of trait space reveals Pareto fronts that constrain adaptation. Nat Ecol Evol. 2019;3(11):1539–1551.

[9] Aggeli D, Li Y, Sherlock G. Changes in the distribution of fitness effects and adaptive mutational spectra following a single first step towards adaptation. Nat Commun. 2021;12(1):5193.