Quantifying microbial fitness in high-throughput experiments

Justus Wilhelm Fink[1,2,*] and Michael Manhart[1,2,3,4,†]

[1] *Institute of Integrative Biology, Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland*
[2] *Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ, USA*
[3] *Department of Environmental Microbiology, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf, Switzerland*
[4] *Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ, USA*

(Dated: August 20, 2024)

Few concepts are as central to evolution as is fitness, and yet the quantification of fitness is often ambiguous. In particular, high-throughput experiments to measure mutant fitness in microbes are increasingly common but vary widely in their definitions of fitness, which makes their results difficult to compare. What are the consequences of these different fitness statistics, and is there a best way to quantify fitness in a given context? Here we systematize the set of possible fitness statistics according to the following three choices: 1) the encoding of relative abundance (e.g., transforming by a log or logit function), 2) the time scale over which to measure the change in relative abundance, and 3) the choice of a reference subpopulation for calculating fitness in bulk competition experiments, such as those using DNA-barcoded mutants. We show that these different choices can lead to significantly different interpretations of mutant fitness, affecting the magnitude of fitness effects, the presence of epistasis, and even the fitness ranking across mutants. This can confound predictions for evolutionary dynamics and gene functions. Altogether our results demonstrate the importance of consistent fitness definitions for reproducible results across experiments.

Keywords: fitness, microbial evolution, competition experiments, gene knockouts, growth curves, DNA barcodes

INTRODUCTION

Fitness describes the fate of mutations as they arise in a population [1] and serves to define other core evolutionary concepts such as trade-offs [2]. At the same time, fitness is an often confusing term in evolutionary biology [3–5]. It is possible to classify fitness measures by their role in theory [6] and their scale of measurement [7–9], but these arguments offer little guidance on how to measure fitness with microorganisms in practice. In particular for the field of microbial ecology and evolution, empirical fitness measurements allow us to test general evolutionary theory with fast-growing model organisms [10–15] and can also be used to detect microbial interactions [16, 17], annotate gene function [18, 19], and understand the spread of antibiotic resistance genes [20, 21].

The classic approach to measuring the fitness effects of mutations uses pairwise competition experiments, where the mutant competes with a wild-type in co-culture [22–24]. This approach is ideal because it closely mimics the dynamics of spontaneous mutations, and thus is typically used in experimental evolution [10, 12, 25–27]. For example, the Long-Term Evolution Experiment (LTEE) has evolved *Escherichia coli* over tens of thousand of generations and measured the fitness of each evolved lineage in competition with the ancestor [11, 26, 28]. Since measuring pairwise competitions becomes infeasible for a large number of strains, a second approach is to measure properties of genotypes in monoculture, like the growth rate, and combine these measurements for a relative fitness estimate. Monoculture growth curves are straightforward to measure, even for entire libraries of single-gene knockout mutants [29–31], but empirical tests have shown that the monoculture growth rate is insufficient to predict the mutant fitness [32–34]. Summary statistics, like the area under the curve (AUC), perform better [35–37]. However, this approach breaks down for strains that engage in cross-feeding or toxin production and any attempts to improve monoculture-based fitness estimates require additional experiments [34, 37] and do not go beyond qualitative fitness rankings [37].

A third and more recent approach is to measure relative fitness in bulk competition experiments, based on high-throughput barcode sequencing [13, 38]. For example, transposon-insertion mutagenesis generates libraries of barcoded gene knockouts that can be tracked as they grow in a single batch culture [39]. Estimating the fitness of each barcode allows us to identify genes that are particularly important in the growth environment [18, 19, 40], in the given genetic background [14, 15, 41, 42], and in the context of an ecological community [16, 17, 42, 43]. Bulk fitness estimation also plays a major role to identify beneficial lineages in experimental evolution [13, 44] (not unlike the tracking of viral pathogens in public health [45–47]) and to test their fitness in new environments [48–52]. Many of these studies make different choices about the design of experiments and the calculation of fitness,

* To whom correspondence should be addressed. Email: justus.fink@caltech.edu; Current address: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

† To whom correspondence should be addressed. Email: mmanhart@rutgers.edu

but without explaining why a certain fitness measure was used or how to compare between data sets.

The inconsistent choices for quantifying microbial fitness make it is difficult to compare fitness across experiments, especially because different groups tend to use different definitions. For example, the LTEE reports relative fitness per generation [11, 26], whereas other evolution experiments exclusively report relative fitness per-cycle [12, 53]. New techniques, like barcode sequencing, often spark their own fitness metrics [18, 40, 54–56]. It is also not clear what the consequences of this choice are; a different fitness metric might lead to a different ranking of mutants in a genetic screen (and the importance we assign to those genes), besides affecting the quantitative fitness value. The probability of fixation for spontaneous mutations [57, 58], the speed of adaptation [59], and the effect size for gene essentiality experiments [18, 19] all depend on our ability to estimate quantitative fitness values. This raises three key questions: How do these fitness statistics differ, or are some equivalent? Can we say which choices are optimal, in the light of first principles or some practical considerations? And does this tell us how we should design these experiments? Previous work has addressed aspects of these questions [6–9, 60] but the arguments are scattered across disciplines and based on generic models of population growth [6, 7, 9, 61], rather than incorporating the specific dynamics of microbes in laboratory experiments. Here we address these questions from a unified framework, using realistic microbial population dynamics with empirical traits to systematically test different fitness metrics.

RESULTS

Predictive power varies across different relative fitness statistics

While there are other notions of fitness relevant in different evolutionary contexts (Sec. S1, Fig. S1), in this article we focus on relative fitness as this is the most common object of high-throughput laboratory experiments in microbes [18, 26, 40, 55, 56]. To generalize the wide range of relative fitness statistics used in these studies, we define a genotype's relative fitness as any number that is sufficient to predict the genotype's relative abundance $x(t)$ over a short-time horizon (Fig. S1). The simplest approach is to take a linear expansion of $x(t)$ and use its slope at time t to predict the change over a time-window Δt:

$$x(t+\Delta t) = x(t) + \left.\frac{dx}{dt}\right|_t \Delta t. \tag{1}$$

Under this linear approximation, the slope $s^{\text{linear}} = dx/dt|_t$ constitutes a relative fitness statistic, since it is sufficient to predict the change in relative abundance. Figure 1A (top panel) shows a schematic trajectory of relative abundance $x(t)$ with two examples of this predictive approach (dashed arrows). As we can see, the naive linear statistic of relative fitness can significantly under- or overestimate the actual change in relative abundance, because $x(t)$ changes nonlinearly in time.

As an alternative, we can transform the relative abundance x into a new variable $m(x)$ that improves the quality of prediction (Fig. 1B; compare [62]). We define an *encoding* as a smooth, strictly-increasing function of relative abundance. This one-to-one mapping allows us to predict the change in relative abundance using a linear expansion of the encoded relative abundance, rather than the relative abundance itself:

$$x(t+\Delta t) = m^{-1}\left(m(x(t)) + s^m \Delta t\right), \tag{2}$$

where m^{-1} is the inverse of the encoding function and

$$s^m = \left.\frac{dm}{dt}\right|_t \tag{3}$$

is the relative fitness under the encoding m. Note that the relative fitness of neutral mutations is always zero, independent of the encoding.

For example, Fig. 1A (middle panel) shows the relative abundance under the log-encoding $m(x) = \log x$; this encoding is implicit in typical plots of relative abundance on a logarithmic scale [44]. This leads to an approximately linear trajectory in the beginning, where the prediction quality of the slope of the encoding $s^{\log}$ is thus higher than for the slope of relative abundance s^{linear} (compare top and middle panels of Fig. 1A). However, under the log-encoding the prediction quality remains poor at later stages in the trajectory. In the case of Fig. 1A, we see that the ideal encoding is given by $m(x) = \text{logit } x$ (bottom panel), where the logit function is defined as

$$\text{logit } x = \log\left(\frac{x}{1-x}\right). \tag{4}$$

This is ideal because it transforms the nonlinear trajectory of relative abundance $x(t)$ into a linear trajectory that can be fully predicted from the linear expansion in Eq. (2).

Why is the logit function $m(x) = \text{logit } x$ the ideal encoding of relative abundance for a fitness statistic in this case? Since we generated the relative abundance $x(t)$ in Fig. 1A using logistic dynamics, the logit function is linearly related to the inverse function of $x(t)$ (Sec. S2). More generally, the ideal encoding for a given relative abundance dynamics $x(t)$ is any linear function of the inverse $m(x(t)) = at(x) + b$ of those dynamics. Note that the ambiguous scale a also includes log and logit encodings with a different base of the logarithm (discussed in Ref. [7]). The inversion of $x(t)$ is equivalent to removing frequency (relative abundance) dependence from the relative fitness statistic s^{linear} [63]. For example, Fig. S2 shows that if we apply the logit-encoding to a trajectory with non-logistic underlying dynamics (e.g., Gompertz

model), the logit transform reduces but does not entirely remove the nonlinearity in the relative abundance trajectory (see the mathematical notes by Mallet [64] for more examples). We note that is possible to generalize the concept of encodings to absolute abundance as well (Sec. S3).

Of course, the exact dynamics of microbial populations are generally not known, and if they were known, there would be no need to make predictions using the linear expansion in Eq. (2). In the absence of an exact model, there are mathematical and practical reasons for why the logit function is a sensible encoding for relative abundance. First, logistic dynamics can be interpreted as the lowest-order approximation to more complex dynamics of relative abundance (Sec. S2). Second, the logit encoding is particularly suited to the binomial sampling noise in experimental measurements of relative abundance (Fig. S3). Logit is the designated link function for binomial random variables, meaning that it normalizes the measurement variance across relative abundances (heteroscedasticity), regardless of whether the underlying dynamics are logistic [65–67]. For these reasons, we exclusively focus on the logit-based relative fitness for the remainder of this work.

Relative fitness statistics require a choice of timescale

In practice, the relative abundance of a genotype is only available as a trajectory of discrete time points (see crossmarks in Fig. S1C). To estimate the relative fitness at a given time point, we calculate the finite difference of the encoded abundance between the current time t and the previous time point:

$$s^m \approx \frac{m(x(t)) - m(x(t-\Delta t)}{\Delta t}, \tag{5}$$

where Δt is the time difference between observations. For discretized time steps it is also possible to define a multiplicative fitness that describes the ratio, rather than the difference, of encoded relative abundance between time points (Sec. S4).

In the case of a single mutant competing against a wild-type strain, the estimated relative fitness of the mutant (using Eq. (5)) under the logit encoding is

$$s^{\text{logit}} = \frac{1}{\Delta t}\left[\log\frac{x_{\text{mut}}(t)}{x_{\text{wt}}(t)} - \log\frac{x_{\text{mut}}(t-\Delta t)}{x_{\text{wt}}(t-\Delta t)}\right], \tag{6}$$

where x_{mut} is the relative abundance of the mutant and $x_{\text{wt}} = 1 - x_{\text{mut}}$ is the relative abundance of the wild-type. This form of relative fitness is widely used in empirical measurements of microbial fitness [12, 66]. Note that a special property of the logit encoding is that we can rewrite Eq. (6) in terms of the log fold-changes of the mutant and wild-type strains:

$$s^{\text{logit}} = \frac{\text{LFC}_{\text{mut}} - \text{LFC}_{\text{wt}}}{\Delta t}, \tag{7}$$

where the log fold-change of each strain is $\text{LFC} = \log N(t)/N(t-\Delta t)$ and N is the biomass of the strain. The LFC is sometimes also referred to as a Malthusian parameter [9, 26, 68].

A key element of this estimate of relative fitness (Eq. (5)) is the time interval Δt. There are three common ways to choose a time interval Δt in empirical measurements of microbial fitness. The simplest is to use a fixed clock time between measurements (e.g., one day). However, many microbial populations, especially in laboratory experiments [26, 69] but also in some natural environments [70–72], grow in discrete growth cycles dictated by pulses of nutrients (batch culture). These cycles define an intrinsic timescale of population dynamics and also determine the time point of sampling. In this case, it is convenient to quantify relative fitness *per cycle*

$$\begin{aligned} s^{\text{logit}}_{\text{cycle}} &= \log\frac{x_{\text{mut}}(t)}{x_{\text{wt}}(t)} - \log\frac{x_{\text{mut}}(t-\Delta t)}{x_{\text{wt}}(t-\Delta t)} \\ &= \text{LFC}_{\text{mut}} - \text{LFC}_{\text{wt}}, \end{aligned} \tag{8}$$

where we have chosen the logit-encoding and $\Delta t = 1$ growth cycle in Eq. (5).

Another important timescale for microbial populations is the generation time. In general, the relative fitness per cycle (Eq. (8)) may depend on the number of generations in the growth cycle, and it can be valuable to normalize for this dependence, especially when comparing across environments [26]. However, defining the number of generations for a population with multiple genotypes growing at different rates is ambiguous. In the case of a single mutant competing with a wild-type, it is common to consider only the number of generations experienced by the wild-type strain, which is estimated by the log fold-change $\text{LFC}_{\text{wt}} = \log N_{\text{wt}}(t)/N_{\text{wt}}(t-\Delta t)$. It is convenient to express the generations in base e to match the natural logarithm in the logit-encoding, but this could equivalently be done by converting all logarithms to base 2. Thus the relative fitness of a mutant *per generation* s^m_{gen} is defined by choosing $\Delta t = \text{LFC}_{\text{wt}}$ in Eq. (5). In the case of the logit-encoding, the relative fitness per-generation is

$$s^{\text{logit}}_{\text{gen}} = \frac{\text{LFC}_{\text{mut}}}{\text{LFC}_{\text{wt}}} - 1, \tag{9}$$

where we have replaced ratios of relative abundances with ratios of absolute abundances in Eq. (6) and rearranged to express completely in terms of log fold-changes (Eq. (7)). The relative fitness per generation $s^{\text{logit}}_{\text{gen}}$ with the logit encoding is equivalent to the fitness statistic $W = 1 + s^{\text{logit}}_{\text{gen}}$ used in the LTEE [26] and other studies [73–77]. Some authors assume a fixed number of generations per growth cycle [12, 15], but since this may not be true, we use the term "per-generation" more strictly for fitness statistics where the wild-type generations are explicitly measured.

Relative fitness per-generation ranks mutants differently than relative fitness per-cycle

Even though relative fitness of a mutant per growth-cycle (Eq. (8)) and per generation (Eq. (9)) are both common statistics in microbial fitness measurements, they are not just equivalent quantities in different units. One major discrepancy between these fitness statistics occurs when the wild-type strain has no growth or experiences net death during the growth cycle because the number of wild-type generations is not well-defined (denominator of Eq. (9)) [5, 78]. This occurs for microbes under high drug concentrations, as well as for microbial populations under harsh environmental conditions as found in sediments or wastewater [79, 80]. Figure 1C shows four qualitatively different scenarios for a mutant competing with a wild-type strain, parameterized according to their LFCs. A comparison between relative fitness per-generation and per-cycle only makes sense when both the wild-type and the mutant show net growth (scenario 1 in Fig. 1C)

However, even when the LFC is positive, the relative fitness per-generation can produce different rankings compared to the relative fitness per-cycle. For example, Fig. 1D shows two mutant genotypes (red and grey) that have opposite rankings under these fitness statistics: the grey mutant has higher relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$, but the red mutant has higher relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$. The disagreement in ranking requires positive covariation in the LFCs (red bow-tie area in Fig. 1D), such that the mutant with higher LFC also induces a higher LFC in the wild-type (Sec. S5, Fig. S4).

Fitness statistics disagree over predicted ranking of single gene knockouts

While we have shown that relative fitnesses per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{gen}}$ can lead to different mutant rankings in principle, we need to know whether this outcome occurs under realistic scenarios of microbial population dynamics. We thus simulate fitness rankings for gene knockout mutants using empirically measured growth traits, based on a previously published growth curve dataset for the single-gene knockout collection in *Saccharomyces cerevisiae* [29, 81]. We estimate growth traits from growth curves (Fig. 2A; Methods; Sec. S6) and find large variation in lag time, growth rate, and biomass yield for the single gene knockouts (grey dots in Fig. 2B,C; also see Fig. S5); despite the knockouts affecting multiple traits (pleiotropy), there are only weak correlations between them. To estimate relative fitness per-cycle (Eq. (8)) and per-generation (Eq. (9)), we simulate a pairwise competition for each knockout against the wild-type (Fig. 2D) using a consumer-resource model with a single limiting nutrient (Methods) [82, 83]. While the relative fitness per-cycle and per-generation are highly correlated across mutants overall (Fig. S6), there are major differences in the ranks of individual mutants (Fig. 2E). For example, measuring the relative fitness per-generation ranks one beneficial mutant (grey dot highlighted with blue circle in Fig. 2E) 145 positions higher than quantifying relative fitness per-cycle (where higher rank corresponds to higher fitness). Indeed, when we plot the set of competitions in this fitness ranking on the LFC diagram, we find that there is many pairs of points that have positive covariation in the mutant and wild-type LFC and thus give rise to ranking differences (compare Figs. 2F and 1D) despite a negative covariation overall (compare Figs. 2F and S4B).

What features of the population dynamics are responsible for the schism between relative fitness per-cycle and per-generation? We explore this by systematically varying the mutant trait distributions and initial conditions in our simulations (Fig. S7). This demonstrates that the fitness statistics $s^{\text{logit}}_{\text{cycle}}$ and $s^{\text{logit}}_{\text{gen}}$ disagree on the ranking if the mutants have diminished biomass yield and are competed at a high initial frequency (50%) against the wild-type (rows A and D in Fig. S7). Intuitively, these conditions are necessary because they allow mutants to alter the wild-type LFC. Even though the wild-type growth rate is unaffected by mutants, the wild-type LFC will change in the presence of mutants that have lower yield because they reduce the time to resource depletion (row D in Fig. S7). A change in wild-type LFC also requires the mutant to be present at high relative abundance initially (row A in Fig. S7; see Sec. S7 for detailed conditions). By this mechanism, it is possible to concoct a set of growth traits where relative fitness per-generation and per-cycle deliver completely opposite rankings (Fig. S8).

As a further example, we test how the choice of fitness statistic affects conclusions from the LTEE. We reanalyze competition data from the LTEE [11] using the relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$ (equivalent to the original definition used in those studies, $W = 1 + s^{\text{logit}}_{\text{gen}}$) and relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ (Sec. S8). We see that the relative fitness statistics disagree on the ranking of some evolved lineages overall (Fig. S9) and at a given time point (Fig. S10). We confirm that the long-term fitness trend in the LTEE is the same for $s^{\text{logit}}_{\text{gen}}$ and $s^{\text{logit}}_{\text{cycle}}$ (Fig. S11), but it is possible to construct hypothetical scenarios of evolution where this would not be the case (Fig. 3A-B). Beyond the qualitative mismatch between these fitness statistics, $s^{\text{logit}}_{\text{cycle}}$ and $s^{\text{logit}}_{\text{gen}}$ can also lead to different conclusions for the presence of epistasis between multiple mutations (Fig. 3C-F,Fig. S12). For example, measuring relative fitness per-generation will detect negative magnitude epistasis between a mutation that only affects lag time and a mutation that only affects biomass yield (Fig. 3C), but measuring relative fitness per-cycle shows no epistasis (Fig. 3D). This is because mutations that increase biomass yield increase the wild-type LFC and, in this case, decrease the relative fitness per-generation (compare Fig. S12A and C). By the same mechanism, relative fitness per-cycle detects

epistasis between mutations that affect growth rate and biomass yield (Fig. 3F), but no such epistasis is present in relative fitness per-generation (Fig. 3E). This again demonstrates that it is possible to arrive at different biological conclusions from the same experimental data depending on the choice of fitness statistic.

Higher-order effects distort relative fitness measured in bulk competitions

So far we have focused on measuring relative fitness in pairwise competition, but this is usually not practical for large numbers of mutants. Measuring traits of genotypes in monocultures to predict their pairwise relative fitness (Sec. S1) is convenient, but when we test this strategy using our simulation framework we find that most traits perform poorly (Sec. S9, Fig. S13) (with some success for the AUC, but this success depends on the choice of time scale; Fig. S14). Another approach to measuring relative fitness is to use bulk competition experiments, where many mutant genotypes compete simultaneously in a single culture and each genotype is tracked through DNA barcode sequencing [39, 44]. However, this raises the question of how well relative fitness of a mutant in bulk competition corresponds to its relative fitness in a pairwise competition with the wild-type, since the growth in bulk might be influenced by the presence of other mutant genotypes, a phenomenon known as a higher-order interaction [83–85].

An important choice in bulk competition experiments is the relative abundance of the library of all mutant genotypes compared to the wild-type. The invasion of a spontaneous mutation into an existing population is best captured by pairwise competitions with low initial relative abundance of the mutant (Case I, Fig. 4A), and one way to recreate this scenario in a bulk competition is to use a low relative abundance for the mutant library overall (Case II, Fig. 4A) [40, 48–52]. A practical problem with Case II is that individual genotypes in the library will have low absolute abundances, which leads to stochasticity in the population dynamics and the sequencing preparation [13, 86, 87]. Therefore, it is common to compete the mutant library by itself (Case III, Fig. 4A) [13, 15, 17–19, 88].

A second important choice is whether to quantify fitness of a mutant relative to the whole population ("total relative fitness"; Fig. 4B, top panel) [18, 19, 88] or relative to another specific genotype, like the wild-type ("pairwise relative fitness"; Fig. 4B, bottom panel; Methods) [40, 48–52]. In the case of a mutant library growing by itself (Case III, Fig. 4A), it is still possible to estimate a pairwise fitness by using known neutral mutants as the reference population [15, 17] and calculating fitness relative to this group (Secs. S10 and S11).

To test the consequences of these choices, we simulate a bulk competition (batch culture) using the trait data for yeast single-gene knockouts and compare the estimates of total and pairwise relative fitness to fitness measured in pairwise competition, which we use as the ground truth (Methods). Figure 4C shows that the total relative fitness has systematically higher error than the pairwise relative fitness does, since the total fitness measures the genotype relative not just to the wild-type but to all other mutants as well (Sec. S12). For example, a mutant that grows identically to the wild-type (neutral phenotype) has zero increase in the pairwise relative abundance, but may have a net increase in the total relative abundance due to the poor growth of other mutants (compare top and bottom panel in Fig. 4B). This affects the total relative fitness of all mutants in a uniform way (Fig. 4C, Sec. S12), such that total and pairwise relative fitness completely agree in the ranking of genotypes (Fig. S15A).

While the pairwise relative fitness is the preferred method for estimating fitness in bulk competitions, the presence of higher-order interactions means it still deviates from the fitness in pairwise competitions. For the specific population dynamics in our simulation (Methods; Sec. S13,), we decompose the higher-order interaction into two terms, a fitness-independent term and a fitness-dependent term that acts as an amplifier for fitness values (Fig. S16; Sec. S14). Intuitively, the population in bulk competition consumes resources more slowly and gives more time for growth rate differences to accrue. The higher-order effects were tested by Levy et al. [13], who compared fitness estimates for ca. 30 beneficial lineages in bulk to the fitness estimates from pairwise competition and found that bulk fitness was typically higher, with some variation from lineage to lineage, consistent with our results here.

Can we reduce these deviations by changing the relative abundance of the mutant library (Fig. 4A)? When we repeat our simulation using a high mutant library abundance (Case III, Fig. 4A), we find that the absolute error for pairwise relative fitness increases proportionally (inset in Fig. 4C). Clearly, minimizing the abundance of the mutant library minimizes the strength of higher-order interactions (also reducing ranking disagreement (Fig. S15B,C).

As low mutant abundances create practical problems for experiments, it is valuable to identify a maximum abundance for the mutant library that keeps the error from bulk fitness estimates below a desired threshold. It is convenient to express this threshold in relative rather than absolute terms (compare Fig. S17 to Fig. 4C). Figure 4D shows the relative error from higher-order interactions across a range of mutant library abundances and using the specific population dynamics of our model, we derive the following rule: Assuming that the mutants in the library only have variation in growth rate, the mutant library abundance should be below

$$x_{\text{lib}} < \epsilon \cdot \frac{g_{\text{wt}}}{|g_{\text{lib}} - g_{\text{wt}}|}, \tag{10}$$

where ϵ is the desired threshold on the relative error,

g_{wt} is the wild-type growth rate, and g_{lib} is the growth rate of the library as a whole (Sec. S15). In the case of our single-gene knockout library, Eq. (10) predicts a maximum library abundance of 24.6% based on the wild-type and library growth rate ($g_{\mathrm{wt}} = 0.406$, $g_{\mathrm{lib}} = 0.389$) for a relative error $\epsilon = 1\%$. Figure 4D shows that this maximum library abundance keeps the relative error below 1% for high-fitness mutants (bright yellow), because they are dominated by growth rate effects, but fails for mutants close to neutrality because they have a trade-off between growth rate and lag time, which the estimate in Eq. (10) neglects. It is possible to derive a more precise bound that keeps the relative error below 1% for all mutant genotypes (here: $x_{\mathrm{lib}} = 0.02\%$, see vertical line in Fig. 4D), but this requires prior knowledge of the trait covariation in the mutant library (Sec. S15).

DISCUSSION

Best practices for quantifying mutant fitness in high-throughput experiments. In this work, we have introduced a conceptual framework (summarized in Fig. 5) to derive common statistics of relative fitness from three essential choices:

1. The choice of the state variable for relative abundance (encoding m; Eq. (2) and Fig. 1B).
2. The choice of the time scale for the change in relative abundance (Δt; Eq. (5)).
3. In bulk competition experiments, the choice of subpopulation that acts as the reference for relative abundance of a chosen mutant genotype (Fig. 4B).

The combination of these choices leads to a range of fitness statistics, including those commonly used in population genetics, experimental evolution, and transposon-insertion screens (Fig. 5). However, as we compare these statistics in simulated competition experiments, we find that the choice of time scale can lead to different mutant rankings (Figs. 2D and 2E) and the choice of the reference subpopulation can lead to a systematic offset in fitness values (Fig. 4C). Based on these insights, we recommend the following choices for quantifying relative fitness of a mutant:

1. Use the logit encoding of relative abundance, because under the null model of logistic dynamics this linearizes the trajectory of relative abundance and regularizes measurement noise (Fig. S3).
2. Use a fixed extrinsic time scale (e.g., a single growth cycle), rather than an intrinsic time scale (e.g., the number of generations) which introduces an additional factor of variation between competitions.
3. In bulk competition experiments, use the pairwise fitness of the mutant relative to the wild-type reference subpopulation (either by including a barcoded wild-type or by grouping neutral mutants into a virtual wild-type) because this more closely matches the relative fitness in pairwise competition.
4. Since relative fitness measurements in bulk competitions inevitably carry an error from higher-order interactions, we recommend minimizing relative abundance of the mutant library as practically feasible (Fig. 4D). For a given error tolerance, Eq. (10) gives an estimate for the maximum library abundance.

If measuring direct competition is not feasible, we empirically find that the area under the growth curve (AUC) is the best approximation of the true fitness ranking (Fig. S13) but one must carefully choose the time scale (Fig. S14). Our recommendations agree with previous criticisms of relative fitness per-generation [5, 8, 9, 78] and total relative fitness [60, 89] but differ from the standard practice in high-throughput genetic screens [18, 19, 88] and many evolution experiments [33, 75], including the LTEE [11, 26]. The choice of the reference group is a well recognized issue in high-throughput evolutionary studies, in which one first estimates a total relative fitness and then subtracts a correction based on the fitness of the wild-type [12, 50, 52] or a mean population fitness [13, 41, 42, 51, 89]. In contrast, we recommend to choose the reference group at the level of relative abundance (Fig. 4B; Eq. (18))

Consequences of fitness quantification choices for microbial ecology and evolution. The quantitative differences between relative fitness statistics affects our ability to make evolutionary predictions. In particular, the fixation probability of a spontaneous mutation depends on the magnitude of relative fitness compared to other timescales like mutation and drift [57, 58], and in microbial populations with clonal interference, it depends on the entire distribution of fitness effects [59, 90]. For example, multiplying the mutation rate with an incompatible measure of fitness would lead to predicting the wrong speed of adaptation. In the context of multiple mutations in the same cell, different fitness statistics can lead to different conclusions about the presence of magnitude epistasis (Fig. 3C-F), and in the context of gene essentiality tests, they may affect the outcome of a significance test (e.g., using log or logit with the test statistic in Ref. [18]).

Using different fitness statistics can even lead to differences in mutant rankings (see Fig. S8 for an extreme case). Since measurements often serve as a first screen to narrow down the investigation to the top set of genes [17, 40], a difference in mutant ranking means the investigation might miss out on relevant genes because of the choice of fitness statistic.

Other sources of discrepancy in quantifying mutant fitness. Besides the conceptual choices of fitness quantification discussed in this article, experimental limitations can create discrepancies between replicate fitness measurements of the same mutant [36, 50, 91]. For example, relative abundance measurements entail sampling uncertainties (when we sample liquid for colony counting or DNA sequencing) [44, 60, 61], copy number variation [18, 44] as well as PCR jackpots and sequencing read errors [13, 86, 87]. Furthermore, there will inevitably

be some variation in initial condition between replicates and fluctuations during the fitness assay [36, 87, 91]. On one hand, the conceptual choices for quantifying fitness can mitigate experimental uncertainty. For example, the logit encoding normalizes the sampling errors over the time series of relative abundance (Fig. S3B). On the other hand, there is sometimes a trade-off between conceptual choices and experimental precision, such as the choice of the initial mutant library abundance, which needs to be low to minimize higher-order interactions and yet high enough to minimize sampling uncertainty [44]. Future work will need to elucidate the interplay between conceptual and experimental sources of discrepancy in fitness measurements.

How general is the disagreement between fitness statistics? The discrepancy between relative fitness per-cycle and per-generation (Figs. 1D and 2E), as well as the offset in the total relative fitness compared to pairwise fitness (Fig. 4C) will hold under any form of population dynamics. However, the details of the population dynamics do matter for the higher-order interactions that cause the discrepancy between relative fitness from bulk and pairwise competitions (Fig. 4D). The empirical evidence for the presence and mechanisms of higher-order interactions in microbial populations is limited [83–85]. Nevertheless, the fact that we see the fact that we see higher-order interactions even under competition for a single resource suggests that these effects will be present under more complex dynamics. Indeed, a comparison of mutant fitness measurements found systematic overestimates for fitness from bulk compared to pairwise competitions [13]. Since this could not be explained by the offset from using total relative fitness, it is indicative of higher-order interactions. Future work should provide more comparisons between relative fitness from bulk and pairwise competitions to better understand higher-order interactions.

METHODS

Inferring growth traits for the single-gene knockout collection in Yeast We use a previously published dataset [29] for the single-gene knockout collection in *Saccharomyces cerevisiae* [81], where the authors track growth of each genotype for 47 hours in monoculture, using microwell plates with defined growth medium. We download this growth curve dataset from the PROPHECY database [92], choosing specifically the dataset measured in Synthetic Defined medium. We last accessed the PROPHECY website (`http://prophecy.lundberg.gu.se/`) in March 30, 2020, but as of May 30, 2024, the website no longer seems accessible and included a reformatted version of the data with our code repository (`https://github.com/justuswfink/24FitnessQuantification`). From the raw timeseries of optical density, the authors subtracted a background correction based on blank wells, applied another correction for nonlinearity at high optical densities and smoothed the growth curve to remove electrical noise [29].

We start with the curves in the published dataset (9951 curves) and apply further trimming and smoothing steps (Sec. S6). From this data we calculate the time series of instantaneous per-capita growth rate $N^{-1}dN/dt$ (where N is the optical density) and identify time windows where the rate is approximately constant, which we interpret as distinct growth phases (Sec. S6). We only include curves that have a single phase of constant exponential growth followed by a stationary phase of approximately zero growth (9424 curves).

We quantify the biomass yield, maximum growth rate, and the lag time from each remaining curve as follows. First, we estimate the initial abundance N_{initial} (average optical density over first three time points) and the final abundance N_{final} (average optical density over stationary phase) from the growth curve. Then we calculate the biomass yield as $Y = (N_{\text{final}} - N_{\text{initial}})/R(0)$, where $R(0) = 111$ mM (20 g/L) is the initial concentration of glucose (assuming glucose is the single limiting resource). To estimate the maximum growth rate, we average the instantaneous growth rate over the exponential phase. Finally, we estimate the lag time from the intersection of the log initial abundance ($\log N_0$) with the slope of the maximum growth rate during the exponential growth phase [93]. We excluded curves with negative initial OD, curves with negative inferred lag times, and curves with a low quality of fit between the measured time series and a simulated curve based on the inferred trait values ($R^2 < 0.95$; see below for the model). Our final database includes trait estimates for 9195 curves, which represents 92.4% of the original dataset [29].

Each single-gene deletion strain was measured in two technical replicate growth curves, using a second plate with identical layout in the same plate reader. Some genotypes have only one estimate of the growth traits because the other replicate did not pass our filters (273 genotypes), but for most genotypes we retain two replicate estimates in our final dataset (4163 genotypes); a few genotypes even have three (2 genotypes) or four replicate estimates (54 genotypes) because these genotypes were included multiple times by the original authors [29]. From these replicates we finally calculate the average yield, growth rate, and lag time for each genotype. The dataset also contained many replicate growth curves of the wild-type strain, 374 of which passed our filters. Since wild-type traits inferred from these replicates had large variation (potentially due to measuring the wild-type across many different plates and days), we define the wild-type trait as the median, not the mean. Although the wild-type measurements included in this dataset have large trait variation (orange dots in Fig. 2B,C), the fact that the variation across knockouts is significantly greater for growth rate (Levene's test $p = 1.3 \times 10^{-9}$) and lag time (Levene's test $p = 0.025$) and that the replicate measurements of gene knockouts

are correlated (Fig. S5) suggests that these trait measurements do capture true genetic variation. In contrast, the variation in biomass yield across knockouts is not significantly different from the variation across wild-type replicates (Levene's test $p = 0.44$), suggesting that the variation in biomass yield is driven by non-genetic factors like the abundance or batch of resources used for each growth measurement.

Simulating population dynamics in competition experiments Throughout this work, we model the competition between a set of genotypes during a single batch culture growth cycle using the following model of population dynamics [82, 83]:

$$\frac{1}{N_j}\frac{dN_j}{dt} = g_j(t, R), \qquad N_j(0) = x_j(0)N_0, \tag{11a}$$

$$\frac{dR}{dt} = -\sum_j \frac{1}{Y_j}\frac{dN_j}{dt}, \qquad R(0) = R_0, \tag{11b}$$

where N_j is the absolute abundance, $x_j(0)$ the initial relative abundance, $g_j(t, R)$ the growth rate, and Y_j the biomass yield of genotype j, where the genotypes include a wild-type as well as one or more mutants. The initial absolute abundance of all genotypes together is N_0, and the concentration of the single limiting resource is R with initial value R_0. Note that Eq. (11b) assumes that cells consume resources only for biomass growth and not for maintenance of existing biomass.

To capture the sigmoidal shape of typical growth curves with a lag phase, exponential growth phase, and saturation phase (e.g., Fig. 2A), we model the dependence of growth rate on time and resource concentration as

$$g_j(t, R) = g_j\Theta(t - \lambda_j)\Theta(R), \tag{12}$$

where Θ is the step function:

$$\Theta(z) = \begin{cases} 0 & z \leq 0, \\ 1 & z > 0. \end{cases} \tag{13}$$

In this model, a genotype has an initial lag phase of time λ_j where no growth occurs, followed by a phase of constant exponential growth at rate g_j, and ending when the resource concentration R reaches zero. The time to resource depletion, which we also call the saturation time, is defined by the implicit equation $R(t_{\text{sat}}) = 0$. The simple form of the growth response (Eq. (12)) means that the absolute abundance of genotype j at saturation is

$$N_j(t_{\text{sat}}) = N_j(t_0)e^{g_j(t_{\text{sat}}-\lambda_j)\Theta(t_{\text{sat}}-\lambda_j)} \tag{14}$$

and its log fold-change is

$$\text{LFC}_j = g_j(t_{\text{sat}} - \lambda_j)\Theta(t_{\text{sat}} - \lambda_j). \tag{15}$$

To calculate relative fitness per-generation (Eq. (9)) or per-cycle (Eq. (8) with $\Delta t = 1$), we numerically determine the saturation time t_{sat} by integrating the biomass and resource dynamics (Eq. (11)) as described in previous work [82, 83]. We note that it also possible to derive an approximate expression for the saturation time t_{sat} which we use for our theoretical calculations (Sec. S7).

Testing total and pairwise relative fitness in bulk competition experiments. We explore total and pairwise relative fitness as two alternative measures of fitness in bulk competition experiments, and defined as follows. Consider a population with multiple genotypes such that each genotype i has absolute abundance $N_i(t)$ at time t. The *total relative abundance* of the genotype is

$$x_i(t) = \frac{N_i(t)}{\sum_j N_j(t)}. \tag{16}$$

The "total" here refers to the fact that this is the abundance of genotype i relative to all other genotypes in the population. Similarly, we can think of the definition of relative fitness discussed in the main text (Eq. 3) as the *total relative fitness* of this genotype under an encoding m:

$$\sigma_i^m = \frac{d}{dt}m(x_i(t)). \tag{17}$$

However, sometimes we want to track the dynamics of a genotype relative to another specific genotype; for example, to follow the fate of a mutant against the wild-type in a bulk competition experiment. For a pair of genotypes $i \neq j$, we thus define the *pairwise relative abundance*

$$x_{ij}(t) = \frac{x_i(t)}{x_i(t) + x_j(t)}, \tag{18}$$

which is the relative abundance of genotype i in a subpopulation of genotype i and j. To predict the change in the pairwise relative abundance, we can use the slope

$$s_{ij}^m = \frac{d}{dt}m(x_{ij}(t)), \tag{19}$$

which defines the *pairwise relative fitness* of genotype i with respect to genotype j. In the special case of a population with only two strains, the pairwise relative fitness and total relative fitness are identical ($\sigma_2^m = s_{21}^m$) but they may differ with more than two genotypes (Sec. S12).

It is also possible to define the total and pairwise relative fitness statistics as finite differences over a time interval, rather than as instantaneous derivatives. For example, we can define them over a growth cycle starting at $t = 0$ and ending at $t = t_{\text{sat}}$:

$$\sigma_{i,\text{cycle}}^m = m(x_i(t_{\text{sat}})) - m(x_i(0)), \tag{20}$$

$$s_{ij,\text{cycle}}^m = m(x_{ij}(t_{\text{sat}})) - m(x_{ij}(0)). \tag{21}$$

For these fitness statistics in the bulk competitions, we use the logit encoding

$$\sigma_{i,\text{cycle}}^{\text{logit}} = \log\frac{x_i(t_{\text{sat}})}{1 - x_i(t_{\text{sat}})} - \log\frac{x_i(0)}{1 - x_i(0)} \tag{22}$$

$$s_{ij,\text{cycle}}^{\text{logit}} = \log\frac{x_i(t_{\text{sat}})}{x_i(0)} - \log\frac{x_j(t_{\text{sat}})}{x_j(0)}, \tag{23}$$

where we have rewritten $s^{\text{logit}}_{ij,\text{cycle}}$ in the form that it is presented in bulk competition experiments [15, 51, 52]. The logit encoding has mathematical advantages for coarse-graining the relative fitness of genotype groups (Sec. S10) but using the log encoding is another common choice in the literature [17, 41, 50]. However, the relative abundances of the individual mutants in our simulated bulk competition experiments are low enough that these two encodings are approximately equivalent ($\text{logit}\, x \approx \log x$ for $x \ll 1$).

Unlike the comparison between per-cycle and per-generation relative fitness where we focused on rank differences (Fig. 2E), here we can evaluate absolute differences in fitness estimates because the total and pairwise fitness in bulk fitness are measured in the same units as the relative fitness in pairwise competition. For all mutant genotypes, we calculate the absolute error between these bulk fitness estimates (Eq. (22),(23)) and the relative fitness in pairwise competitions (Eq. (6)), which we take as the ground truth (Fig. 4C). Note that the pairwise competition could depend on the initial relative abundance of the mutant; we have chosen a very low relative abundance (10^{-6}) that mimics a mutant arising de novo and where this dependence is very weak.

ACKNOWLEDGMENTS

JWF wishes to thank Luis-Miguel Chevin, Olivier Tenaillon, Henrique Teotónio and the participants of the 2021 ENS autumn course on Experimental Evolution for valuable discussions early on. Both authors are grateful to Benjamin Raach and Gatwa Tshinsele-Van Bellingen for a critical reading of the manuscript. JWF and MM were supported by an Ambizione grant from the Swiss National Science Foundation (PZ00P3_180147).

[1] John H. Gillespie. *Population Genetics: A Concise Guide*. JHU Press, 2004.
[2] S. C. Stearns. Trade-Offs in Life-History Evolution. *Functional Ecology*, 3(3):259–268, 1989.
[3] G. De Jong. The Fitness of Fitness Concepts and the Description of Natural Selection. *The Quarterly Review of Biology*, 69(1):3–29, 1994.
[4] André Ariew and R. C. Lewontin. The Confusions of Fitness. *The British Journal for the Philosophy of Science*, 55(2):347–363, 2004.
[5] Hanna Kokko. The stagnation paradox: the ever-improving but (more or less) stationary population fitness. *Proceedings of the Royal Society B: Biological Sciences*, 288(1963):20212145, 2021.
[6] Daniel Jefferson Smith, Guilhem Doulcier, Pierrick Bourrat, et al. Why there are so many definitions of fitness in models. *EcoEvoRxiv*, 10.32942/X2V61T, 2024.
[7] Chevin Luis-Miguel. On measuring selection in experimental evolution. *Biology Letters*, 7(2):210–213, 2011.
[8] David Houle, Christophe Pélabon, Günter P. Wagner, and Thomas F. Hansen. Measurement and Meaning in Biology. *The Quarterly Review of Biology*, 86(1):3–34, 2011.
[9] Günter P. Wagner. The Measurement Theory of Fitness. *Evolution*, 64(5):1358–1376, 2010.
[10] Charlotte E. Paquin and Julian Adams. Relative fitness can decrease in evolving asexual populations of S. cerevisiae. *Nature*, 306(5941):368–371, 1983.
[11] M. J. Wiser, N. Ribeck, and R. E. Lenski. Long-Term Dynamics of Adaptation in Asexual Populations. *Science*, 342(6164):1364–1367, 2013.
[12] Sergey Kryazhimskiy, Daniel P. Rice, Elizabeth R. Jerison, and Michael M. Desai. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, 344(6191):1519–1522, 2014.
[13] Sasha F. Levy, Jamie R. Blundell, Sandeep Venkataram, et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181–186, 2015.
[14] Alex N. Nguyen Ba, Ivana Cvijović, José I. Rojas Echenique, et al. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature*, 575(7783):494–499, 2019.
[15] Alejandro Couce, Anurag Limdi, Melanie Magnan, et al. Changing fitness effects of mutations through long-term bacterial evolution. *Science*, 383(6681):eadd1417, 2024.
[16] Jeremy E. Schreier, Christa B. Smith, Thomas R. Ioerger, and Mary Ann Moran. A mutant fitness assay identifies bacterial interactions in a model ocean hot spot. *Proceedings of the National Academy of Sciences*, 120(12):e2217200120, 2023.
[17] Emily C. Pierce, Manon Morin, Jessica C. Little, et al. Bacterial–fungal interactions revealed by genome-wide analysis of bacterial mutant fitness. *Nat Microbiol*, 6(1):87–102, 2021.
[18] Kelly M. Wetmore, Morgan N. Price, Robert J. Waters, et al. Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *mBio*, 6(3):e00306–15, 2015.
[19] Morgan N. Price, Kelly M. Wetmore, R. Jordan Waters, et al. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*, 557(7706):503–509, 2018.
[20] Dan I Andersson. The biological cost of mutational antibiotic resistance: any practical conclusions? *Current Opinion in Microbiology*, 9(5):461–465, 2006.
[21] Alvaro San Millan and R. Craig MacLean. Fitness Costs of Plasmids: a Limit to Plasmid Transmission. *Microbiology Spectrum*, 5(5):5.5.02, 2017.
[22] Santiago F. Elena, Lynette Ekunwe, Neerja Hajela, et al. Distribution of fitness effects caused by random insertion mutations in Escherichia coli. *Genetica*, 102(0):349, 1998.
[23] Clifford Zeyl and J Arjan G M DeVisser. Estimates of the Rate and Distribution of Fitness Effects of Spontaneous Mutation in Saccharomyces cerevisiae. *Genetics*, 157(1):53–61, 2001.
[24] Isabel Gordo, Lilia Perfeito, and Ana Sousa. Fitness Effects of Mutations in Bacteria. *Journal of Molecular Microbiology and Biotechnology*, 21(1-2):20–35, 2012.

[25] Daniel E. Dykhuizen. Experimental Studies of Natural Selection in Bacteria. *Annual Review of Ecology and Systematics*, 21(1):373–398, 1990.

[26] Richard E. Lenski, Michael R. Rose, Suzanne C. Simpson, and Scott C. Tadler. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist*, 138(6):1315–1341, 1991.

[27] Andrew J Spiers, Sophie G Kahn, John Bohannon, et al. Adaptive Divergence in Experimental Populations of Pseudomonas fluorescens. I. Genetic and Phenotypic Bases of Wrinkly Spreader Fitness. *Genetics*, 161(1):33–46, 2002.

[28] Jeffrey E. Barrick, Zachary D. Blount, Devin M. Lake, et al. Daily Transfers, Archiving Populations, and Measuring Fitness in the Long-Term Evolution Experiment with Escherichia coli. *JoVE (Journal of Visualized Experiments)*, (198):e65342, 2023.

[29] Jonas Warringer, Elke Ericson, Luciano Fernandez, et al. High-resolution yeast phenomics resolves different physiological features in the saline response. *PNAS*, 100(26):15724–15729, 2003.

[30] Guillaume Chevereau, Marta Dravecká, Tugce Batur, et al. Quantifying the Determinants of Evolutionary Dynamics Leading to Drug Resistance. *PLOS Biology*, 13(11):e1002299, 2015.

[31] Manuel Campos, Sander K Govers, Irnov Irnov, et al. Genomewide phenotypic analysis of growth, cell morphogenesis, and cell cycle events in Escherichia coli. *Molecular Systems Biology*, 14(6):e7573, 2018.

[32] Jeniffer Concepción-Acevedo, Howard N. Weiss, Waqas Nasir Chaudhry, and Bruce R. Levin. Malthusian Parameters as Estimators of the Fitness of Microbes: A Cautionary Tale about the Low Side of High Throughput. *PLOS ONE*, 10(6):e0126915, 2015.

[33] Paulo Durão, Massimo Amicone, Lília Perfeito, and Isabel Gordo. Competition dynamics in long-term propagations of Schizosaccharomyces pombe strain communities. *Ecology and Evolution*, 11(21):15085–15097, 2021.

[34] Yoav Ram, Eynat Dellus-Gur, Maayan Bibi, et al. Predicting microbial growth in a mixed culture from growth curve data. *Proceedings of the National Academy of Sciences*, 116(29):14698–14707, 2019.

[35] Kathleen Sprouffske and Andreas Wagner. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, 17(1):172, 2016.

[36] Sarah B. Worthan, Robert D. P. McCarthy, and Megan G. Behringer. Case Studies in the Assessment of Microbial Fitness: Seemingly Subtle Changes Can Have Major Effects on Phenotypic Outcomes. *Journal of Molecular Evolution*, 91:311—-324, 2023.

[37] Désirée A Schmitz, Tobias Wechsler, Ingrid Mignot, and Rolf Kümmerli. Predicting bacterial interaction outcomes from monoculture growth and supernatant assays. *ISME Communications*, 4(1):ycae045, 2024.

[38] Milo S. Johnson, Sandeep Venkataram, and Sergey Kryazhimskiy. Best Practices in Designing, Sequencing, and Identifying Random DNA Barcodes. *Journal of Molecular Evolution*, 91:263—-280, 2023.

[39] Tim van Opijnen and Andrew Camilli. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol*, 11(7):435–442, 2013.

[40] Bidong D. Nguyen, Miguelangel Cuenca V., Johannes Hartl, et al. Import of Aspartate and Malate by DcuABC Drives H2/Fumarate Respiration to Promote Initial Salmonella Gut-Lumen Colonization in Mice. *Cell Host & Microbe*, 27(6):922–936.e6, 2020.

[41] Milo S. Johnson, Alena Martsul, Sergey Kryazhimskiy, and Michael M. Desai. Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science*, 366(6464):490–493, 2019.

[42] Joao A. Ascensao, Kelly M. Wetmore, Benjamin H. Good, et al. Quantifying the local adaptive landscape of a nascent bacterial community. *Nature Communications*, 14(1):248, 2023.

[43] Jonathan N V Martinson, Jeremy M Chacón, Brian A Smith, et al. Mutualism reduces the severity of gene disruptions in predictable ways across microbial communities. *The ISME Journal*, 17(12):2270–2278, 2023.

[44] Jamie R. Blundell and Sasha F. Levy. Beyond genome sequencing: Lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. *Genomics*, 104(6):417–430, 2014.

[45] Nicholas G. Davies, Sam Abbott, Rosanna C. Barnard, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, 372(6538):eabg3055, 2021.

[46] Muhammad Saqib Sohail, Raymond H Y Louie, Zhenchen Hong, et al. Inferring Epistasis from Genetic Time-series Data. *Molecular Biology and Evolution*, 39(10):msac199, 2022.

[47] Matthijs Meijers, Denis Ruchnewitz, Jan Eberhardt, et al. Population immunity predicts evolutionary trajectories of SARS-CoV-2. *Cell*, 186(23):5151–5164.e13, 2023.

[48] Sandeep Venkataram, Barbara Dunn, Yuping Li, et al. Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell*, 166(6):1585–1596.e22, 2016.

[49] Yuping Li, Dmitri A. Petrov, and Gavin Sherlock. Single nucleotide mapping of trait space reveals Pareto fronts that constrain adaptation. *Nature Ecology & Evolution*, 3(11):1539–1551, 2019.

[50] Grant Kinsler, Kerry Geiler-Samerotte, and Dmitri A Petrov. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife*, 9:e61271, 2020.

[51] Christopher W Bakerlee, Angela M Phillips, Alex N Nguyen Ba, and Michael M Desai. Dynamics and variability in the pleiotropic effects of adaptation in laboratory budding yeast populations. *eLife*, 10:e70918, 2021.

[52] Kara Schmidlin, Sam Apodaca, Daphne Newell, et al. Distinguishing mutants that resist drugs via different mechanisms by examining fitness tradeoffs. *eLife*, 13:RP94144, 2024.

[53] Michael J. McDonald, Tim F. Cooper, Hubertus J. E. Beaumont, and Paul B. Rainey. The distribution of fitness effects of new beneficial mutations in *Pseudomonas fluorescens*. *Biology Letters*, 7(1):98–100, 2011.

[54] Chuan Li, Wenfeng Qian, Calum J. Maclean, and Jianzhi Zhang. The fitness landscape of a tRNA gene. *Science*, 352(6287):837–840, 2016.

[55] Pascal M. Frey, Julian Baer, Judith Bergada-Pijuan, et al. Quantifying Variation in Bacterial Reproductive Fitness: a High-Throughput Method. *mSystems*, 6(1):e01323–20, 2021.

[56] Tzahi Gabzi, Yitzhak Pilpel, and Tamar Friedlander. Fitness Landscape Analysis of a tRNA Gene Reveals that the Wild Type Allele is Sub-optimal, Yet Mutationally Robust. *Molecular Biology and Evolution*, 39(9):msac178, 2022.

[57] Motoo Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713–719, 1962.

[58] Z Patwa and L.M Wahl. The fixation probability of beneficial mutations. *Journal of The Royal Society Interface*, 5(28):1279–1289, 2008.

[59] Michael M Desai and Daniel S Fisher. Beneficial Mutation–Selection Balance and the Effect of Linkage on Positive Selection. *Genetics*, 176(3):1759–1798, 2007.

[60] Sebastian Matuszewski, Marcel E Hildebrandt, Ana-Hermina Ghenu, et al. A Statistical Guide to the Design of Deep Mutational Scanning Experiments. *Genetics*, 204(1):77–87, 2016.

[61] Anurag Limdi and Michael Baym. Resolving Deleterious and Near-Neutral Effects Requires Different Pooled Fitness Assay Designs. *Journal of Molecular Evolution*, 91(3):325–333, 2023.

[62] C Denniston and J F Crow. Alternative fitness models with the same allele frequency dynamics. *Genetics*, 125(1):201–205, 1990.

[63] B. R. Levin, J. Antonovics, H. Sharma, et al. Frequency-dependent selection in bacterial populations. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 319(1196):459–472, 1997.

[64] James Mallet. The struggle for existence. how the notion of carrying capacity obscures the links between demography, darwinian evolution and speciation, 2012.

[65] P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Routledge, Boca Raton, 2 edition, 1989.

[66] Romain Gallet, Tim F. Cooper, Santiago F. Elena, and Thomas Lenormand. Measuring Selection Coefficients Below 10^{-3}: Method, Questions, and Prospects. *Genetics*, 190(1):175–186, 2012.

[67] Jeff Gill and Michelle Torres. *Generalized Linear Models*. Sage Publications, Thousand Oaks, 2 edition, 2020.

[68] Helen K. Alexander, Guillaume Martin, Oliver Y. Martin, and Sebastian Bonhoeffer. Evolutionary rescue: linking theory for conservation and medicine. *Evolutionary Applications*, 7(10):1161–1179, 2014.

[69] Vaughn S. Cooper. Experimental Evolution as a High-Throughput Screen for Genetic Adaptations. *mSphere*, 3(3):e00121–18, 2018.

[70] Christoph A. Thaiss, David Zeevi, Maayan Levy, et al. Transkingdom Control of Microbiota Diurnal Oscillations Promotes Metabolic Homeostasis. *Cell*, 159(3):514–529, 2014.

[71] Antonio M. Martin-Platero, Brian Cleary, Kathryn Kauffman, et al. High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nature Communications*, 9(1):266, 2018.

[72] Daniel Muratore, Angela K. Boysen, Matthew J. Harke, et al. Complex marine microbial communities partition metabolism of scarce resources over the diel cycle. *Nature Ecology & Evolution*, 6(2):218–229, 2022.

[73] Javier DelaFuente, Jeronimo Rodriguez-Beltran, and Alvaro San Millan. Methods to Study Fitness and Compensatory Adaptation in Plasmid-Carrying Bacteria. In Fernando De La Cruz, editor, *Horizontal Gene Transfer*, volume 2075, pages 371–382. Springer US, New York, NY, 2020.

[74] Autumn Peterson, Carina Baskett, William C. Ratcliff, and Anthony Burnetti. Using light for energy: examining the evolution of phototrophic metabolism through synthetic construction, 2022.

[75] R. Z. Moger-Reischer, J. I. Glass, K. S. Wise, et al. Evolution of a minimal cell. *Nature*, 620(7972):122–127, 2023.

[76] Ellen Baake, Adrián González Casanova, Sebastian Probst, and Anton Wakolbinger. Modelling and simulating Lenski's long-term evolution experiment. *Theoretical Population Biology*, 127:58–74, 2019.

[77] Sivan Pearl Mizrahi, Akshit Goyal, and Jeff Gore. Community interactions drive the evolution of antibiotic tolerance in bacteria. *Proceedings of the National Academy of Sciences*, 120(3):e2209043119, 2023.

[78] R. E. Lenski. On The Use Of Selection Rate Versus Relative Fitness. `https://lenski.mmg.msu.edu/ecoli/srvsrf.html`. Accessed: 2021-06-25.

[79] B.J. Koch, T.A. McHugh, M. Hayer, et al. Estimating taxon-specific population dynamics in diverse microbial communities. *Ecosphere*, 9(1):e02090, 2018.

[80] Chunxiao Wang, Yu Yang, Yulin Wang, et al. Absolute quantification and genome-centric analyses elucidate the dynamics of microbial populations in anaerobic digesters. *Water Research*, 224:119049, 2022.

[81] Guri Giaever, Angela M. Chu, Li Ni, et al. Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, 418(6896):387–391, 2002.

[82] Michael Manhart, Bharat V. Adkar, and Eugene I. Shakhnovich. Trade-offs between microbial growth phases lead to frequency-dependent and non-transitive selection. *Proceedings of the Royal Society B: Biological Sciences*, 285(1872):20172459, 2018.

[83] Michael Manhart and Eugene I. Shakhnovich. Growth tradeoffs produce complex microbial communities on a single limiting resource. *Nature Communications*, 9(1):1–9, 2018.

[84] Jonathan Friedman, Logan M. Higgins, and Jeff Gore. Community structure follows simple assembly rules in microbial microcosms. *Nature Ecology & Evolution*, 1(5):1–7, 2017.

[85] Manon A. Morin, Anneliese J. Morrison, Michael J. Harms, and Rachel J. Dutton. Higher-order interactions shape microbial interactions as microbial community complexity increases. *Scientific Reports*, 12(1):22640, 2022.

[86] Ryan Seamus McGee, Grant Kinsler, Dmitri Petrov, and Mikhail Tikhonov. Improving the Accuracy of Bulk Fitness Assays by Correcting Barcode Processing Biases. *Molecular Biology and Evolution*, 41(8):msae152, 2024.

[87] Grant Kinsler, Kara Schmidlin, Daphne Newell, et al. Extreme Sensitivity of Fitness to Environmental Conditions: Lessons from #1BigBatch. *Journal of Molecular Evolution*, 91(3):293–310, 2023.

[88] Manon Morin, Emily C Pierce, and Rachel J Dutton. Changes in the genetic requirements for microbial interactions with increasing community complexity. *eLife*, 7:e37072, 2018.

[89] Fangfei Li, Marc L. Salit, and Sasha F. Levy. Unbiased Fitness Estimation of Pooled Barcode or Amplicon Sequencing Studies. *Cell Systems*, 7(5):521–525.e4, 2018.

[90] Benjamin H. Good, Michael J. McDonald, Jeffrey E. Barrick, et al. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678):45–50, 2017.

[91] Andrew J. Borchert, Alissa Bleem, and Gregg T. Beckham. Experimental and Analytical Approaches for Improving the Resolution of Randomly Barcoded Transposon Insertion Sequencing (RB-TnSeq) Studies. *ACS Synthetic Biology*, 11(6):2015–2021, 2022.

[92] Luciano Fernandez-Ricaud, Jonas Warringer, Elke Ericson, et al. PROPHECY—a database for high-resolution phenomics. *Nucleic Acids Research*, 33(suppl_1):D369–D373, 2005.

[93] M. H. Zwietering, I. Jongenburger, F. M. Rombouts, and K. van 't Riet. Modeling of the Bacterial Growth Curve. *Applied and Environmental Microbiology*, 56(6):1875–1881, 1990.

[94] Farida Vasi, Michael Travisano, and Richard E. Lenski. Long-Term Experimental Evolution in Escherichia coli. II. Changes in Life-History Traits During Adaptation to a Seasonal Environment. *The American Naturalist*, 144(3):432–456, 1994.

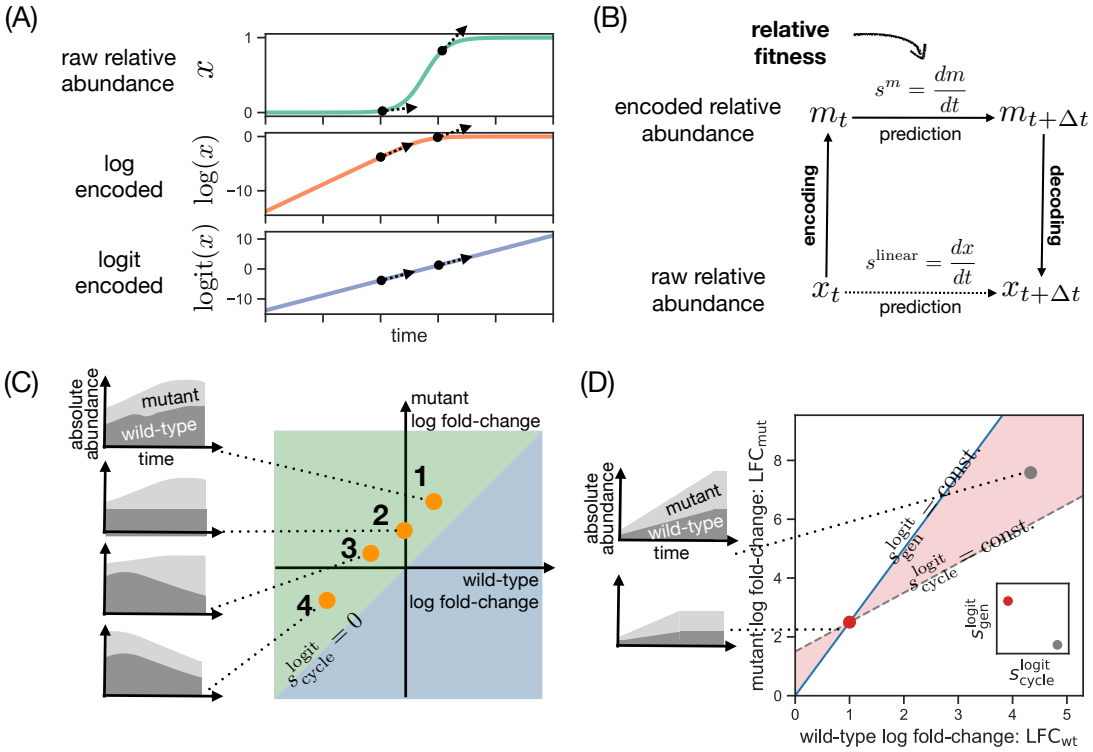

FIG. 1. **Overview of the choice of encoding and the choice of timescale for quantifying relative fitness.** (A) Example trajectory of relative abundance x (upper panel) for a mutant invading and eventually replacing a wild-type population. The same trajectory is plotted under the encoding $\log x$ (middle panel) and the logit-encoding $\log(x/(1-x))$ (lower panel). (B) The general flow-chart to predict the future relative abundance of a mutant given a relative fitness value $s^m = dm/dt$ for some encoding m. The current relative abundance x_t is transformed into the new variable $m_t = m(x_t)$, then projected into the future through a linear extrapolation using s^m (upper horizontal arrow) and finally converted back into a frequency $x_{t+\Delta t}$ using the decoding function m^{-1}. (C) Four scenarios for positive mutant fitness with different underlying population dynamics. For each scenario, we show an example trajectory of absolute abundance (stacked) for the wild-type (dark grey) and mutant population (light grey). Each scenario is mapped as a single-dot onto the fold-change diagram (center plot) and colored areas indicated positive (green area) and negative relative fitness per-cycle (blue area; compare Eq. (8)). (D) Basic constellation for misranking between relative fitness per-cycle $s^{\mathrm{logit}}_{\mathrm{cycle}}$ and relative fitness per-generation $s^{\mathrm{logit}}_{\mathrm{gen}}$. For a given competition (red dot), misranking occurs in a bow-tie area of the fold-change space (red shade). Any competition in the right half of this area (grey dot) will have higher mutant fitness $s^{\mathrm{logit}}_{\mathrm{cycle}}$ but lower mutant fitness $s^{\mathrm{logit}}_{\mathrm{gen}}$ (right inset). As small plots on the left, we show possible population dynamics that generate this fold-change variation.

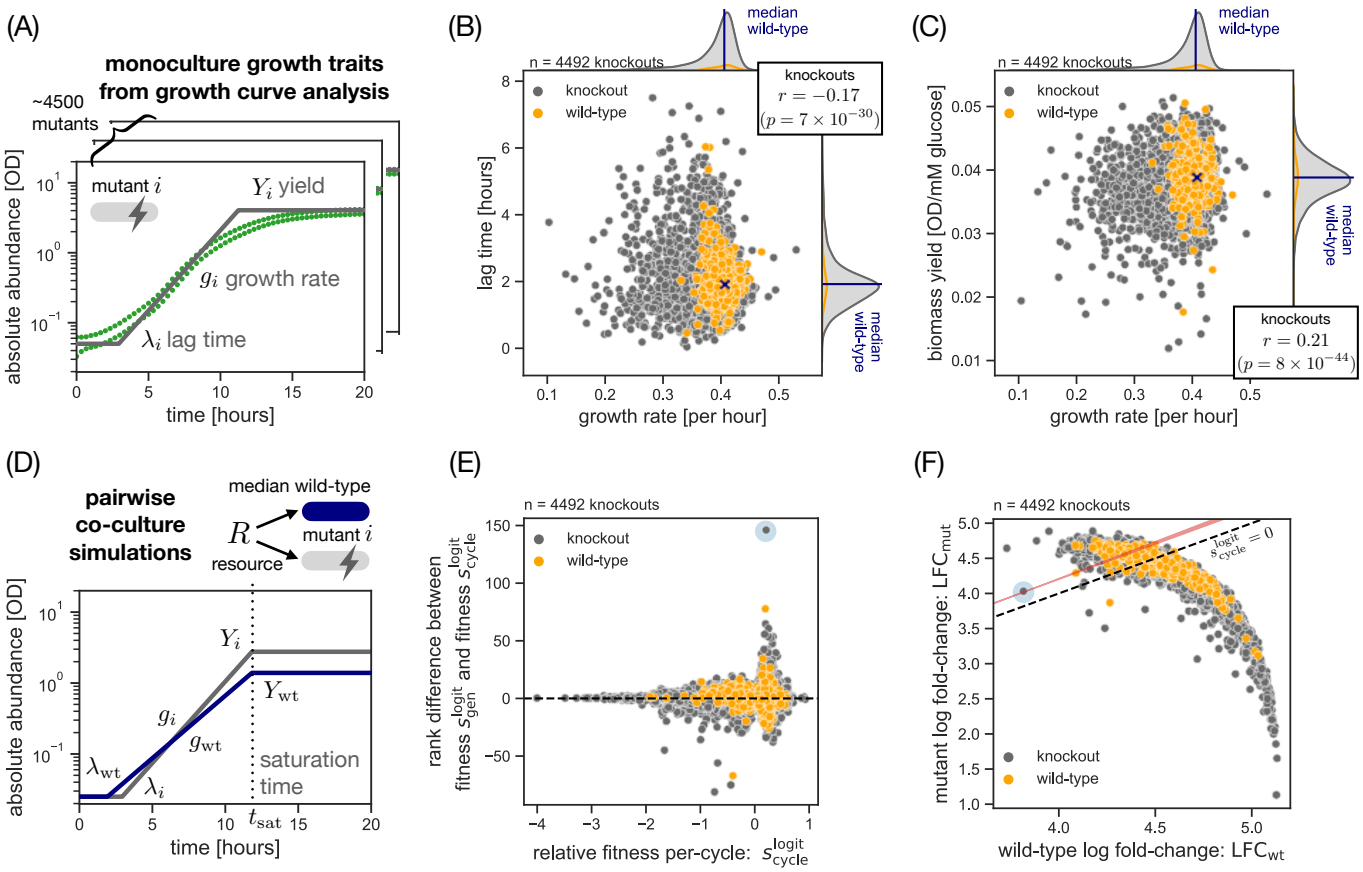

FIG. 2. **Comparison of mutant fitness rankings with different statistics on empirical trait variation.** (A) Overview of the growth curve dataset and the estimated growth traits for the knockout library of *Saccharomyces cerevisiae* (Methods). (B) Covariation between estimated steady-state growth rate g and lag time λ across all mutant strains (grey dots; Pearson correlation coefficient $r = -0.17$, $p = 7 \times 10^{-30}$) as well as wild-type replicates (orange dots; $r = -0.16$, $p = 0.002$). The reference wild-type strain for our pairwise co-culture simulations is defined by the median trait values (black cross) of all wild-type replicate. (C) Covariation between measured steady-state growth rate g and biomass yield Y across all mutant strains (grey dots; $r = 0.21$, $p = 8 \times 10^{-44}$) as well as wild-type replicates (orange dots; $r = -0.06$, $p = 0.25$). (D) Overview of pairwise co-culture simulations. For each mutant strain (orange), we simulate a competition growth cycle against a reference wild-type strain using the estimated traits (panel A) and laboratory parameters for the initial condition ($N_0 = 0.05$ OD, $R_0 = 111$ mM glucose, $x = 0.5$; Methods) and quantify relative fitness of the mutant in different statistics (Eq. (8),Eq. (9)). (E) Rank disagreement between relative-fitness per-generation $s^{\text{logit}}_{\text{gen}}$ and per-cycle $s^{\text{logit}}_{\text{cycle}}$. For each fitness statistic, we calculate the mutant ranking (higher rank means higher fitness and mutants with equal fitness are assigned the lowest rank in the group). The rank difference is defined as the rank in $s^{\text{logit}}_{\text{gen}}$ minus the rank in $s^{\text{logit}}_{\text{cycle}}$. (F) Covariation between wild-type and mutant fold-change across all simulated competitions, with mutant strains (grey dots) and wild-type replicates (orange dots). For each wild-type replicate, we simulate a pairwise co-culture competition against the reference wild-type strain. We highlight the mutant with the greatest rank difference (blue halo) in panel E and F, and its corresponding bow-tie area of misranking (compare Fig. 1D).

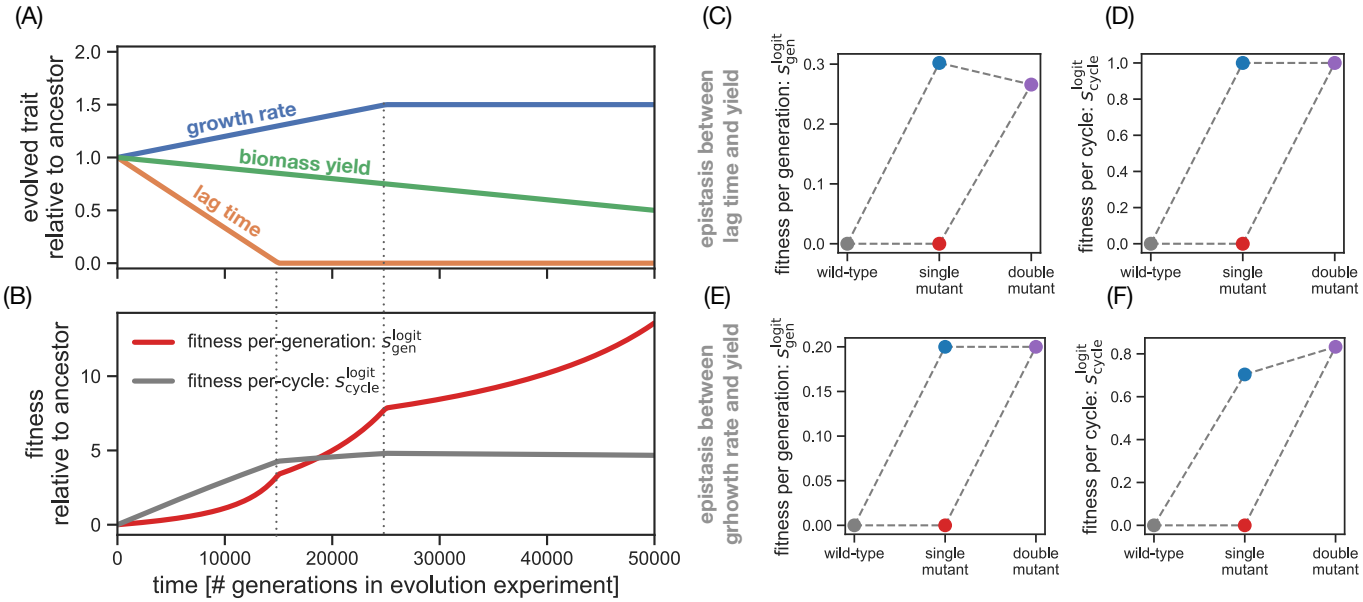

FIG. 3. **Potential consequences of the choice of fitness statistic for the interpretation of evolutionary data.** (A) Hypothetical scenario for the trait evolution in a long-term evolution experiment. An evolving population decreases in lag time (orange line), increases in growth rate to a maximum (blue line) and keeps decreasing in biomass yield (green line). This trend is similar to initial observations from the LTEE [94]. (B) Corresponding long-term trend in relative fitness based on the trait evolution in panel A. We estimate relative fitness per-cycle $s_{\text{cycle}}^{\text{logit}}$ (grey line) and per-generation $s_{\text{gen}}^{\text{logit}}$ (red line) every 250 generations. Dotted grey lines mark the end of trait evolution in lag time and growth rate in panel A. For the actual fitness trend in the LTEE, see Fig. S11. (C) Epistasis plot for lag time and yield using relative fitness per-generation $s_{\text{gen}}^{\text{logit}}$. Colored dots show the fitness for a single mutant with shorter lag time (blue dot), a single mutant with higher biomass yield (red dot) and a double mutant with both mutations (purple dot). (D) Epistasis plot for lag time and yield in relative fitness per-cycle $s_{\text{cycle}}^{\text{logit}}$. (E) Epistasis plot for growth rate and yield in relative fitness per-generation $s_{\text{gen}}^{\text{logit}}$. Colored dots show the fitness for a single mutant with higher growth rate (blue dot), a single mutant with higher biomass yield (red dot) and a double mutant with both mutations (purple dot). (F) Epistasis plot for growth rate and yield in relative fitness per-cycle $s_{\text{cycle}}^{\text{logit}}$. All epistais plots are based on 50:50 competition growth cycles with the wild-type (compare panels C,D and Fig. S12A-C, compare panels E,F and Fig. S12D-F).

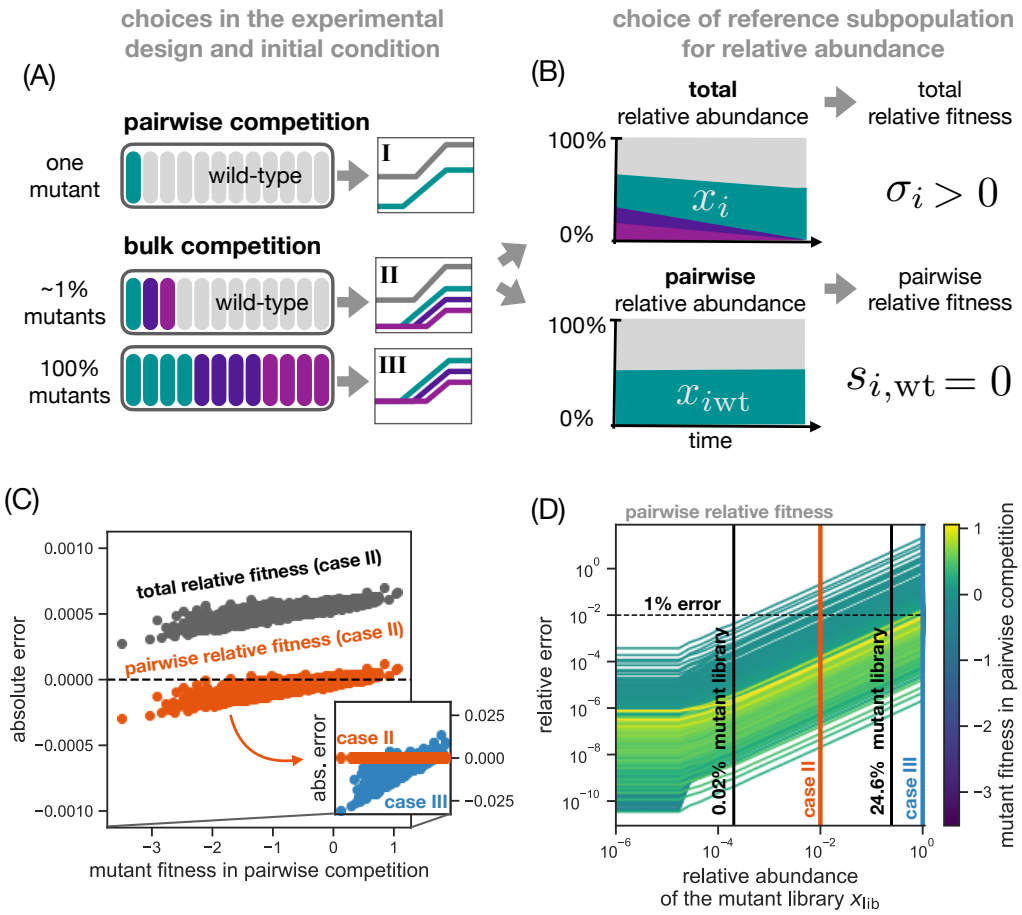

FIG. 4. **The choice of library abundance and reference group in bulk competition experiments.** (A) Overview of a pairwise competition experiment (upper row) and multiple scenarios for bulk competition experiments (middle and bottom row) with different initial fraction of the mutant library (colored ovals) in the inoculum (open box). For each scenario, we show a schematic growth cycle (log absolute abundance) in the inset on the right. (B) Schematic relative abundance trajectories for a mutant compared to two alternative subpopulations. We distinguish between the total relative abundance x_i with respect to the population as a whole (height of green band in the top box) and the pairwise relative abundance $x_{i\mathrm{wt}}$ with respect to the wild-type (height of green band in the bottom box; Eq. (18). We indicate the sign of total relative fitness (Eq. (22)) and pairwise relative fitness (Eq. (23)) on the right. (C) The absolute error between bulk and pairwise competition experiments. The total relative fitness (grey dots; Eq. (22)) and the pairwise relative fitness (red dots; Eq. (23)) for empirical knockouts (Fig. 2A) in bulk competition growth cycle with low mutant library abundance (panel A, case II; Methods). The absolute error is defined as the bulk fitness statistic minus the relative fitness in pairwise competition (Eq. (S81)). In the inset, the absolute error for pairwise relative fitness (Eq. (23)) for a bulk competition growth cycle with high mutant library abundance (blue dots; case III). The x-axis and the red dots in the inset are identical to the main plot. (D) The relative error in bulk competition experiments as a function of mutant library abundance in the inoculum. Each line corresponds to a knockout in our dataset, and represents the relative error between the pairwise relative fitness in bulk competition and the relative fitness in pairwise competition (Eq. (S91)). In black lines, we show the recommended mutant library abundance for our dataset based on Eq. (10) ($x_{\mathrm{lib}} \approx 24.6\%$) and based on Eq. (S109) ($x_{\mathrm{lib}} \approx 0.02\%$, Sec. S15).

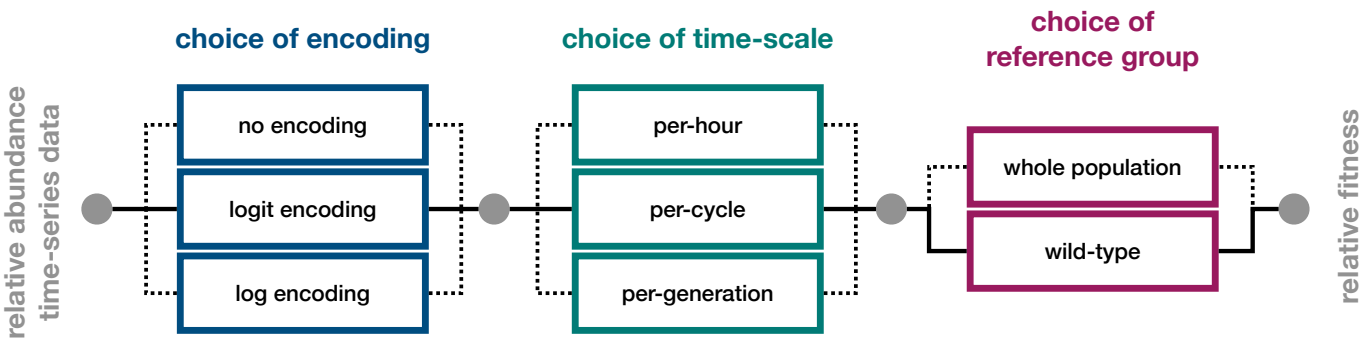

FIG. 5. **Flow-diagram for the quantification of relative fitness from time-series data.** Given the relative abundance of the mutant genotype at two consecutive timepoints (e.g. the start and end of a growth cycle), the user has to choose an encoding (Fig. 1A-B) and the time-scale for evaluating the change in relative abundance (Eq. (5)). In bulk competition experiments, multiple definitions of relative abundance are possible depending on the choice of the reference subpopulation (compare Fig. 4B; Methods). Each combination of these choices (dashed black lines) leads to a different fitness statistic and we summarize our recommendations (thick black line) in the discussion.

Supplementary Information: Quantifying microbial fitness in high-throughput experiments

Justus Wilhelm Fink[1,2,*] and Michael Manhart[1,2,3,4,†]

[1] *Institute of Integrative Biology, Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland*
[2] *Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ, USA*
[3] *Department of Environmental Microbiology, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Dübendorf, Switzerland*
[4] *Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ, USA*

(Dated: August 20, 2024)

S1. DIFFERENT TYPES OF FITNESS UNDER EXAMPLE MODELS OF POPULATION DYNAMICS

We distinguish between three related but distinct notions of fitness [1], all based on predicting dynamics of a population [2–5]. The first type of fitness is *absolute fitness*, which is a property of a single genotype by itself and serves to predict the change in the genotype's absolute abundance $N(t)$ over a future time window Δt (Fig. S1A). This is important for questions about extinction and evolutionary rescue [6]. The second type of fitness is *relative fitness*, which is a property of two genotypes as it describes how the relative abundance $x(t)$ of one genotype changes compared to the other over a time Δt (Fig. S1B). This is important to determine the fixation probability of new mutations [7, 8]. In general these dynamics are stochastic [7, 9, 10], but throughout this paper we focus on their average behavior across replicate cultures (as sketched in Fig. S1C).

A practical challenge of working with relative fitnesses is that they must be measured between all pairs of genotypes in co-culture competitions. Therefore it is common to infer relative fitness of two genotypes based on some individual properties of the genotypes [11–14]. We denote this third notion of fitness as the *fitness potential*; it is a property of an individual genotype, but unlike the absolute fitness, it has no meaning by itself; it is the ratio or difference of fitness potentials that is used to derive relative fitness between two genotypes [15, 16]. The collection of fitness potential values across a large set of genotypes forms a fitness landscape [17, 18]. We note that fitness as defined here gives information about short-time dynamics but not necessarily the long-term outcome (compare [19]). For example, this excludes the ratio of growth rates [14, 20] or the resource concentration R^* in chemostat equilibrium [21] since these quantities cannot tell you how fast the absolute or relative abundance is changing.

In this section we explicitly calculate relative fitness of a mutant under a few example models of population dynamics, using the different encodings of relative abundance as described in the main text (Fig. 1A). Consider a competition coculture between a wild-type genotype with absolute abundance $N_{\text{wt}}(t)$ and a mutant genotype with absolute abundance $N_{\text{mut}}(t)$. We can describe their dynamics according to the ordinary differential equations (ODEs)

$$\begin{aligned} \frac{1}{N_{\text{wt}}}\frac{dN_{\text{wt}}}{dt} &= g_{\text{wt}}(t, N_{\text{wt}}, N_{\text{mut}}), \\ \frac{1}{N_{\text{mut}}}\frac{dN_{\text{mut}}}{dt} &= g_{\text{mut}}(t, N_{\text{wt}}, N_{\text{mut}}). \end{aligned} \tag{S1}$$

Note that the per-capita growth rates g_{wt} and g_{mut} of each genotype can depend on both genotypes to reflect competition or other interactions. The relative abundance of the mutant genotype at time t is

$$x(t) = \frac{N_{\text{mut}}(t)}{N_{\text{mut}}(t) + N_{\text{wt}}(t)}. \tag{S2}$$

The dynamics of the mutant relative abundance are therefore described by

$$\frac{dx}{dt} = \Big[g_{\text{mut}}(t, N_{\text{wt}}, N_{\text{mut}}) - g_{\text{wt}}(t, N_{\text{wt}}, N_{\text{mut}})\Big] \times x(1-x). \tag{S3}$$

As defined in the main text (Eq. (3)), the relative fitness of a mutant is $s^m = dm/dt$ for an encoding $m(x)$ of the relative abundance x. Under the trivial linear encoding ($m(x) = x$), the relative fitness is therefore just the right-hand side of the relative abundance ODE (Eq. (S3)):

$$\begin{aligned} s^{\text{linear}} &= \frac{dx}{dt} \\ &= \Big[g_{\text{mut}}(t, N_{\text{wt}}, N_{\text{mut}}) - g_{\text{wt}}(t, N_{\text{wt}}, N_{\text{mut}})\Big] \times \\ &\quad x(1-x). \end{aligned} \tag{S4}$$

* To whom correspondence should be addressed. Email: justus.fink@caltech.edu; Current address: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

† To whom correspondence should be addressed. Email: mmanhart@rutgers.edu

For the log encoding $m(x) = \log x$, we use the identity $d\log x/dt = x^{-1}dx/dt$ to obtain the log-encoded relative fitness:

$$\begin{aligned} s^{\log} =& \frac{d}{dt}\log x \\ =& \Big[g_{\text{mut}}(t, N_{\text{wt}}, N_{\text{mut}}) - g_{\text{wt}}(t, N_{\text{wt}}, N_{\text{mut}})\Big] \times \\ & (1-x). \end{aligned} \tag{S5}$$

Finally, for the logit encoding $m(x) = \operatorname{logit} x = \log(x/(1-x))$, we use the identity $d\operatorname{logit} x/dt = x^{-1}(1-x)^{-1}dx/dt$ to obtain the logit-encoded relative fitness:

$$\begin{aligned} s^{\text{logit}} =& \frac{d}{dt}\operatorname{logit} x \\ =& g_{\text{mut}}(t, N_{\text{wt}}, N_{\text{mut}}) - g_{\text{wt}}(t, N_{\text{wt}}, N_{\text{mut}}). \end{aligned} \tag{S6}$$

By comparing relative fitness values under the linear encoding (Eq. (S4)) and under the logit encoding (Eq. (S6)), we see how the logit encoding has removed the explicit dependence on the mutant relative abundance (factors of x and $1-x$), although there can be implicit dependence on the mutant relative abundance within the per-capita growth rates of each strain (g_{wt} and g_{mut}) due to density-dependent growth rates.

If the per-capita growth rates g_{wt} and g_{mut} (Eq. (S1)) are constants, then these constant growth rates also act as fitness potentials since they each depend only on a single genotype but their difference determines relative fitness (under the logit encoding, Eq. (S6)) between the genotypes. The growth rate is not a fitness potential under more complex dynamics, however. For example, consider a competition model with explicit density dependence:

$$\begin{aligned} \frac{1}{N_{\text{wt}}}\frac{dN_{\text{wt}}}{dt} &= r_{\text{wt}}\left(1 - \frac{N_{\text{wt}}}{K_{\text{wt}}} - \frac{N_{\text{mut}}}{K_{\text{mut}}}\right), \\ \frac{1}{N_{\text{mut}}}\frac{dN_{\text{mut}}}{dt} &= r_{\text{mut}}\left(1 - \frac{N_{\text{wt}}}{K_{\text{wt}}} - \frac{N_{\text{mut}}}{K_{\text{mut}}}\right), \end{aligned} \tag{S7}$$

where the growth rates decrease as the genotype abundances reach their carrying capacities K_{wt} and K_{mut}, and the maximum growth rates at low abundances are r_{wt} and r_{mut}. In this case, the relative fitness under the logit encoding is (from Eq. (S6))

$$s^{\text{logit}} = (r_{\text{mut}} - r_{\text{wt}})\left(1 - \frac{N_{\text{wt}}}{K_{\text{wt}}} - \frac{N_{\text{mut}}}{K_{\text{mut}}}\right). \tag{S8}$$

In this case, there is no fitness potential because it is not possible to separate Eq. (S8) into a difference between terms that only depend on each genotype separately.

S2. THE ROLE OF LOGISTIC POPULATION DYNAMICS IN LOGIT-ENCODED RELATIVE FITNESS

Here we show how the logit encoding of relative abundance is related to the logistic model of population dynamics. For a relative abundance x, logistic dynamics are

$$\frac{dx}{dt} = rx(1-x), \tag{S9}$$

where r is the exponential rate at which relative abundance increases from low values. This form emerges from the general dynamics of relative abundance (Eq. (S3)) when the difference in per-capita growth rates g_{mut} and g_{wt} is constant. Once can also interpret the logistic model as a lowest-order approximation for more complex dynamics. That is, consider a general equation for relative abundance:

$$\frac{dx}{dt} = f(x), \tag{S10}$$

for an arbitrary function $f(x)$. Since this function must obey the boundary conditions $f(0) = 0$ and $f(1) = 0$ (the relative abundance must stop changing when it either goes extinct or fixes), a polynomial expansion of $f(x)$ must have roots at these values:

$$f(x) = rx(1-x)\left(1 + \mathcal{O}(x)\right). \tag{S11}$$

Thus the logistic model in Eq. (S9) is a lowest-order approximation even when the true dynamics are more complex.

The logistic differential equation in Eq. (S9) has the solution

$$x(t) = \frac{x(0)e^{rt}}{x(0)e^{rt} + 1 - x(0)}. \tag{S12}$$

The logit encoding of the logistic relative abundance has linear dependence on time:

$$\begin{aligned} m(x) &= \operatorname{logit} x \\ &= rt + \log\left(\frac{1-x(0)}{x(0)}\right). \end{aligned} \tag{S13}$$

This is another way to see why the relative fitness under the logit encoding (the time derivative of Eq. (S13)) is constant under logistic dynamics. Mathematically, this occurs because the logit function is the inverse of the logistic dynamics (Eq. (S12)), up to a shift and rescaling. Thus if the relative abundance dynamics are different from logistic (Eq. (S9)), the logit encoding no longer exactly linearizes the trajectory of relative abundance and thus is no longer the optimal encoding for relative fitness (see example in Fig. S2).

S3. DEFINITION OF ABSOLUTE FITNESS FOR A GENOTYPE

Here we give an explicit definition of a genotype's absolute fitness, analogous to the definition of relative fitness in the main text (Eqs. (1)–(3)). Conceptually, absolute

fitness is any number that is sufficient to predict a genotype's absolute abundance N over a short time window. Let an encoding $m(N)$ be any smooth, strictly-increasing function of the absolute abundance N. We can then predict the absolute abundance over a time window Δt using a linear expansion of the encoded abundance (analogous to main text Eq. (2) for relative fitness):

$$N(t+\Delta t) \approx m^{-1}\left(m(N(t)) + a^m \Delta t\right), \tag{S14}$$

where m^{-1} is the inverse of the encoding function and

$$a^m = \frac{dm}{dt} \tag{S15}$$

is defined as the absolute fitness of the genotype under the encoding m (analogous to main text Eq. (3) for relative fitness; see also Fig. 1B). For example, the absolute fitness of a genotype under the log encoding $m(N) = \log N$ is the per-capita growth rate:

$$a^{\log} = \frac{1}{N}\frac{dN}{dt}. \tag{S16}$$

In general, the ideal encoding of absolute abundance is the inverse function of the absolute abundance trajectory $N(t)$ (up to a shift and rescaling), so that the first-order expansion in Eq. (S14) is exact and the absolute fitness a^m is sufficient to determine changes in absolute abundance up to any future time. The log encoding $m(N) = \log N$ is therefore ideal when absolute abundance grows or decays exponentially, while the logit encoding $m(N) = \operatorname{logit} N$ is ideal for a population that grows with logistic density dependence (Eq. (S7) in case of a single genotype).

Absolute fitness and relative fitness are related, since the relative abundance of a genotype is determined by normalizing its absolute abundance by the absolute abundance of all genotypes in the population. Specifically, the relative fitness of a genotype is determined by the absolute fitnesses for all genotypes in the population. For example, in the case of two genotypes, the relative fitness under the logit encoding (Eq. (S6)) is the difference between the genotypes' absolute fitness under the log encoding (Eq. (S16)). In the case of constant per-capita growth rates, these log-encoded absolute fitnesses also act as fitness potentials.

S4. RELATIVE FITNESS PREDICTIONS IN DISCRETE TIME: ADDITIVE VS. MULTIPLICATIVE FORM

In our framework, relative fitness is a statistic that predicts relative abundance in an additive equation (Eq. (2)) but sometimes the dynamics of individual genotypes are modeled using a multiplicative form of fitness. In this section, we show how these multiplicative fitness statistics are related to the additive fitness statistics, in particular for the logit encoding.

We consider a population of a wild-type and a mutant genotype, where we track the mutant's relative abundance over multiple rounds of competition (e.g. growth cycles)

$$x(0) \to x(1) \to \cdots \to x(r-1) \to x(r) \tag{S17}$$

and the variable x changes each round according to some underlying population dynamics.

In a modeling approach typical to many studies in population genetics [22], these population dynamics are captured in the genotype-specific growth factors

$$f_{\text{mut}} = \frac{N_{\text{mut}}(r+1)}{N_{\text{mut}}(r)}, \qquad f_{\text{wt}} = \frac{N_{\text{wt}}(r+1)}{N_{\text{wt}}(r)}, \tag{S18}$$

which drive the update equation for the mutant relative abundance

$$x(r+1) = \frac{x(r) f_{\text{mut}}}{(1-x(r)) f_{\text{wt}} + x(r) f_{\text{mut}}}, \tag{S19}$$

which is the discrete-time analogue to a differential equation (Eq. (S1)). We divide Eq. (S19) by $1 - x(r+1)$ to obtain the form

$$\frac{x(r+1)}{1-x(r+1)} = \frac{x(r)}{1-x(r)} \frac{f_{\text{mut}}}{f_{\text{wt}}}, \tag{S20}$$

which allows us to recognize the ratio of growth factors $f_{\text{mut}}/f_{\text{wt}}$ as a relative fitness statistic, since it is sufficient to predict the relative abundance of the mutant genotype (under the encoding $m(x) = x/1-x$). But the statistic $f_{\text{mut}}/f_{\text{wt}}$ acts as a multiplying factor in Eq. (S20), whereas the general form of relative fitness s^m (Eq. (2)) acts as an additive factor. What's the relationship between the multiplicative and the additive form of discrete-time relative fitness statistics?

To describe discrete rounds of population dynamics in our framework, we treat the relative abundance $x(r)$ as samples from a continuous timeseries, separated by a time-gap $\Delta t = t(r+1) - t(r)$. For a chosen encoding m, The relative abundance of the mutant genotype in the future round is predicted by

$$m(x(r+1)) = m(x(r)) + s^m \Delta t \tag{S21}$$

since this is how we defined relative fitness s^m in Eq. (2) This relative fitness acts as additive factor, but we apply the exponential function on both sides of Eq. (S21) to obtain the updated equation

$$\exp\left[m(x(r+1))\right] = \exp\left[m(x(r))\right] \cdot \exp(s^m \Delta t), \tag{S22}$$

where the additive fitness s^{logit} together with the timescale Δt acts as a multiplicative factor. Specifically for the logit encoding $m(x) = \operatorname{logit} x$ we have

$$\frac{x(r+1)}{1-x(r+1)} = \frac{x(r)}{1-x(r)} \exp(s^{\text{logit}} \Delta t). \tag{S23}$$

We compare Eq. (S23) to Eq. (S20) and solve for the relative fitness of the mutant genotype

$$s^{\text{logit}} = \frac{\log f_{\text{mut}} - \log f_{\text{wt}}}{\Delta t}, \tag{S24}$$

as a function of the growth factors $f_{\text{mut}}, f_{\text{wt}}$. Using the definition of the growth factors (Eq. (S18)), we see that Eq. (S24) is simply the discrete-time relative fitness for the logit encoding in terms of the mutant and wild-type LFC (Eq. (S30)). As a general point, we note that the growth-factor f_{mut} qualifies as an absolute fitness (since it is sufficient to predict the absolute abundance $N(r+1)$), but does not constitute a relative fitness statistic (since we also need to know f_{wt}, see Eq.(S20) or Eq. (S24)).

More generally, for a given encoding function $m(x)$ we define the mutant's multiplicative relative fitness over a discrete growth cycle as

$$w^m = \exp(s^m \Delta t) \tag{S25}$$

where s^m is the additive relative fitness for this growth-cycle under the chosen encoding (Eq. (3)) and Δt is the duration of the growth-cycle in time units that match s^m. Equation (S25) shows that the additive fitness s^m has time units, but the multiplicative fitness does not and is formally a dimensionless quantity. These dimensionless units are preserved in the approximate formula

$$w^m \approx 1 + s^m \Delta t, \tag{S26}$$

which is the first-order expansion of Eq. (S25) in the limit of weak selection ($|s^m \Delta t| \ll 1$). For example, if we choose to measure relative fitness with the logit encoding on the time-scale per-cycle ($\Delta t = 1$ cycle), we get

$$w^{\text{logit}} \approx 1 + s^{\text{logit}}_{\text{cycle}} \cdot 1 \text{ cycle} \tag{S27}$$

which is the same multiplicative fitness as if we choose to measure the relative fitness on the time-scale per-generation ($\Delta t = \text{LFC}^{\text{wt}}$):

$$w^{\text{logit}} \approx 1 + s^{\text{logit}}_{\text{gen}} \text{LFC}^{\text{wt}}. \tag{S28}$$

In both cases, the time-units for the discrete-time additive fitness cancel in the product term.

So how would a mutant ranking in the multiplicative fitness w^{logit} rank a set of mutant genotypes compared to the fitness statistics $s^{\text{logit}}_{\text{cycle}}$ or $s^{\text{logit}}_{\text{gen}}$? Equation (S27) shows that the multiplicative fitness statistic w^{logit} agrees with the relative fitness per-cycle s^{logit}, but can differ from the ranking in the relative fitness per-generation $s^{\text{cycle}}_{\text{gen}}$ as both terms of the product in Eq. (S28) depend on the mutant (compare Fig. 2F and Sec. S7).

Finally, we address the question how the relative fitness $s^{\text{logit}}_{\text{cycle}}$ agrees in the ranking with the multiplicative fitness w^{logit}, but disagrees with the fitness statistic defined in the Long-Term Evolution Experiment [23] as

$$W = 1 + s^{\text{logit}}_{\text{gen}}. \tag{S29}$$

By comparing Eq. (S29) to the multiplicative fitness in Eq. (S26), we see that the LTEE fitness statistic W does not derive from the general form of the multiplicative fitness (Eq. (S25)) as all statistics derived from this approximation have a multiplying factor Δt that cancels the units of time in s^m. The fact that the LTEE fitness statistic W (Eq. (S29)) misses the term LFC_{wt} compared to the logit-based multiplicative fitness w^{logit} (Eq. (S28)) means that the objects have different units of time, and different rankings.

S5. DERIVATION OF MISMATCH CONDITIONS FOR RELATIVE FITNESS PER-CYCLE AND PER-GENERATION

In this section, we derive the conditions for a ranking mismatch between the relative fitness per-cycle and per-generation (for the logit-encoded relative abundance) across a set of competition experiments. Consider a batch culture where a competing wild-type and mutant genotype have log fold-changes LFC_{wt} and LFC_{mut} over a single growth cycle. The LFCs are convenient variables to describe these dynamics since we can express the mutant's relative fitness per-cycle as (main text Eq. (8))

$$s^{\text{logit}}_{\text{cycle}} = \text{LFC}_{\text{mut}} - \text{LFC}_{\text{wt}} \tag{S30}$$

and the mutant's relative fitness per-generation as (main text Eq. (9))

$$s^{\text{logit}}_{\text{gen}} = \frac{\text{LFC}_{\text{mut}}}{\text{LFC}_{\text{wt}}} - 1. \tag{S31}$$

Here we assume both LFCs are nonzero (so that measuring fitness per-generation is meaningful; see discussion in main text).

To determine how these two fitness statistics lead to different rankings, we consider two competition experiments A and B, which may represent two different mutants competing against the same wild-type or the same mutant tested in two different environments. A mismatch in ranking occurs when the fitness per-cycle in competition is highest in B, while the fitness per-generation is highest in A:

$$\begin{aligned} s^{\text{logit}}_{\text{cycle}}(\text{LFC}^B_{\text{wt}}, \text{LFC}^B_{\text{mut}}) &> s^{\text{logit}}_{\text{cycle}}(\text{LFC}^A_{\text{wt}}, \text{LFC}^A_{\text{mut}}), \\ s^{\text{logit}}_{\text{gen}}(\text{LFC}^B_{\text{wt}}, \text{LFC}^B_{\text{mut}}) &< s^{\text{logit}}_{\text{gen}}(\text{LFC}^A_{\text{wt}}, \text{LFC}^A_{\text{mut}}), \end{aligned} \tag{S32}$$

We insert the expressions for relative fitness per-cycle (Eq. (S30)) and per-generation (Eq. (S31)) into Eq. (S32) to rewrite the condition for ranking mismatch in terms of the LFCs:

$$\text{LFC}^B_{\text{mut}} > \text{LFC}^B_{\text{wt}} + (\text{LFC}^A_{\text{mut}} - \text{LFC}^A_{\text{wt}}), \tag{S33a}$$

$$\text{LFC}^B_{\text{mut}} < \text{LFC}^B_{\text{wt}} \cdot \frac{\text{LFC}^A_{\text{mut}}}{\text{LFC}^A_{\text{wt}}}. \tag{S33b}$$

For a given competition A, Eq. (S33) defines an area in the space of LFCs where competition B can lie such that competitions' fitness is ranked differently per-cycle versus per-generation (Fig. 1D shows an example as the red-shaded area). Biologically, these constraints describe a situation where both the mutant and wild-type LFCs are higher in competition B than in competition A (i.e., so that they gray point is up and to the right of the red point in Fig. 1D), but the LFC increases must be sufficiently balanced between the mutant and wild-type (i.e., so that the point lies within the red area in Fig. 1D).

Typically, however, the LFCs of the wild-type and the mutant are not independent, since these LFCs are jointly constrained by the fact both strains compete for the same finite resources. For example, assume that a single limiting resource with concentration R is consumed in proportion to the growth of each genotype's biomass according to

$$\frac{dR}{dt} = -\frac{1}{Y_{\text{wt}}}\frac{dN_{\text{wt}}}{dt} - \frac{1}{Y_{\text{mut}}}\frac{dN_{\text{mut}}}{dt}, \tag{S34}$$

where Y_{wt} and Y_{mut} are the wild-type and mutant biomass yields (stoichiometry of biomass to resource). We can integrate Eq. (S34) to obtain

$$R(t) = R(0) - \frac{N_{\text{wt}}(t) - N_{\text{wt}}(0)}{Y_{\text{wt}}} - \frac{N_{\text{mut}}(t) - N_{\text{mut}}(0)}{Y_{\text{mut}}}. \tag{S35}$$

The growth cycle stops when no resource remains ($R(t) = 0$), such that

$$R(0) + \frac{N_{\text{wt}}(0)}{Y_{\text{wt}}} + \frac{N_{\text{mut}}(0)}{Y_{\text{mut}}} = \frac{N_{\text{wt}}(0)e^{\text{LFC}_{\text{wt}}}}{Y_{\text{wt}}} + \frac{N_{\text{mut}}(0)e^{\text{LFC}_{\text{mut}}}}{Y_{\text{mut}}}, \tag{S36}$$

where we have expressed the genotype abundances at the end of the growth cycle in terms of their LFCs. Equation (S36) thus entails a constraint between the wild-type and mutant LFCs. For a set of a mutant competitions with the same initial resource concentration $R(0)$, initial abundances $N_{\text{wt}}(0)$ and $N_{\text{mut}}(0)$, and yields Y_{wt} and Y_{mut}, the mutant and wild-type LFCs are constrained by Eq. (S36) to fall along a one-dimensional curve (black line in Fig. S4A). Geometrically, we see that this constraint on LFCs is incompatible with the requirements for a ranking mismatch between fitness per-cycle and per-generation for a pair of mutant competitions (compare black line and red shaded areas in Fig. S4A). However, if some mutant competitions deviate from this constraint, for example by having different yields Y_{mut} or initial conditions $N_{\text{wt}}(0)$ and $N_{\text{mut}}(0)$, then ranking mismatches may be possible (Fig. S4B).

S6. ANALYSIS OF GROWTH CURVES TO IDENTIFY GROWTH PHASES

In this section, we describe in more detail how we identify growth phases from the original dataset of growth curves and use this to choose a subset of curves that matches the simplified growth dynamics of our population dynamics model (Fig. 2A). As mentioned in the main text (Methods), we downloaded this original data from the PROPHECY DATABASE (`http://prophecy.lundberg.gu.se/`), downloading specifically the dataset for growth in Synthetic Defined medium as first analysed and reported in [24]. The original growth curve data is already corrected for background and instrument nonlinearities [24] (summarized in Methods), but we decided to apply additional corrections as follows: To begin with, we concatenate the original 51 data files (for different plate reader runs) into a single, consecutive dataframe and manually handle a duplication in one of the files (Experiment NO. 18). This file has no measurements for the first timepoint ($t = 0$) due to technical error in the original data export [24] and for curves from this experimental run, we set the initial time point to `NAN` value in Python, meaning that these points will be ignored for any subsequent calculations of averages. More generally, we decided to trim the first four timepoints of all growth curves (equivalent to 1h20min from 47h total) and remove OD measurement below a noise threshold (OD = 0.001) as this improves the quality of the fit later on.

After pre-processing, we estimate a smooth time series for the instantaneous growth rate in each growth curve, using a previously published script `gaussianprocess.py` by Swain et al. [25] that implements the Gaussian Process approach to smoothing (download from `https://swainlab.bio.ed.ac.uk/software.html`). We apply this script to the logarithmic absolute abundance $\log \text{OD}$, and reconstruct a smoothed trajectory $f(t) \approx \log \text{OD}(t)$ as well as the first derivative df/dt and the second derivative d^2f/d^2t [25]. Effectively, the script estimates three hyperparameters that capture the shape of each curve and we find that the estimation works best if we constrain the parameter ranges as outline in Table. S1.

symbol	range	interpretation
`th[0]`	$[-5, 5]$	amplitude of variation on log scale
`th[1]`	$[-5, 5]$	flexibility on log scale
`th[2]`	$[-3.83, 0.478]$	measurement error on log scale

TABLE S1. Parameter settings for Gaussian Process optimisation.

With a smoothed time series at hand, we now identify 'plateaus' of constant growth rate using the functions available in Scipy [26]. For each growth curve, we start by identifying so-called 'plateau seeds', which are small intervals where the second derivative is below a chosen threshold $d^2f/d^2t < 5 \cdot 10^{-6}$. For each 'plateau seed' k in

the growth curve, we calculate the average growth rate $\hat{g}_k$ in the time-window of the plateau. Due to some noise in the second derivative, we find many 'plateau seeds' that are adjacent and need to be merged. To do so, we iterate over the 'plateau seeds' in the growth curve and merge the current candidate k with the previous plateau $k-1$, except one of the following conditions is true:

- Both plateaus seeds have a duration that is too long (equal or greater than 100 minutes).
- The transition time between the plateau seeds is too large (equal or greater than 200 minutes).
- The first plateau has a significantly different growth rate that the second one, sucht that the following equation is satisfied

$$|g_{k-1} - g_k| > 0.001 + |g_k| \cdot 0.4, \tag{S37}$$

  where g_k, g_{k-1} are the average growth rate in each plateau (per minute).

Empirically, we find that the duration of these merged plateaus is shorter than what one would expect from visually inspecting the growth curve. Therefore we extend the remaining plateaus in each growth curve as follows: For each plateau, we estimate the lower and upper growth rate df/dt in the time window and take this as a growth rate corridor. We extend the plateau to the left, until df/dt leaves that growth rate corridor, and similarly extend to the right. By definition, the resulting plateau is equal or larger to the original time-window and we recal the average growth rate over the time window.

From this analysis, we obtain a list of growth phases for each curve that allows us to choose a subset of curves that match our model of population dynamics (Methods). We only choose curves that have two plateaus, where the first plateau has significant growth (exponential phase), and the second plateau has no growth (stationary phase). Here we define significant growth as the average growth rate in the plateau time window is larger or equal to `0.0011` per minute. This forms the set of growth curves that we use to estimate growth traits (9424 curves).

S7. THE SATURATION TIME IN OUR MODEL OF POPULATION DYNAMICS

In this section, we restate an explicit expression for the saturation time t_{sat} in pairwise competition that was derived in earlier work and allows us to see how mutants can influence resource depletion and the wild-type LFC (Eq. (15)). Using the same model of population dynamics (Methods, Eq(11)), previous work [27, 28] derived an approximate formula for t_{sat} that shows how it depends on the underlying parameters:

$$t_{\text{sat}} \approx \bar{\lambda} + \bar{\tau}\text{LFC}, \tag{S38}$$

where

$$\bar{\lambda} = \bar{\tau}\bar{Y} \sum_j \frac{x_j g_j \lambda_j}{Y_j} \tag{S39}$$

is the effective lag time of the population,

$$\bar{\tau} = \left[\sum_j x_j g_j \frac{\bar{Y}}{Y_j} \right]^{-1} \tag{S40}$$

is the effective e-fold growth time (reciprocal exponential growth rate) of the population,

$$\bar{Y} = \left[\sum_j \frac{x_j}{Y_j} \right]^{-1} \tag{S41}$$

is the effective yield of the population, and

$$\text{LFC} = \log\left(1 + \frac{R_0 \bar{Y}}{N_0}\right) \tag{S42}$$

is the log fold-change of the total biomass, which depends on the initial absolute abundance N_0 of all genotypes and the initial concentration of resources R_0.

Equation (S38) shows how each competing genotype influences the saturation time t_{sat} and thus the LFCs of all other genotypes via Eq. (15). For example, adding a mutant with slow growth rate increases the effective doubling time (Eq. (S40)), while a mutant with long lag time will increase the effective lag time (Eq. (S39)). Genotypes also influence the saturation time (Eq. (S38)) through the effective yield $\bar{Y}$, which is the harmonic average of yields for all genotypes (Eq. (S41)). The harmonic average means that adding mutants with low biomass yield Y_j can significantly shorten the saturation time, but mutants that are more efficient (high Y_j) have little influence on the duration of the growth cycle. Note that genotypes must be at sufficiently high relative abundance to significantly influence the effective population traits, since each the contribution of each genotype j is weighted by its relative abundance x_j.

S8. ANALYSIS OF FITNESS TRAJECTORIES FROM THE LONG-TERM EVOLUTION EXPERIMENT

A previous analysis of the Long-Term Evolution Experiment (LTEE) performed by Wiser et al.[29] found that evolved populations of *Escherichia coli* increased in relative fitness over 50,000 generations without converging to a maximum fitness. Here we re-analyze the same data by directly comparing the relative fitness (under the logit encoding) per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{gen}}$ to see if the choice of fitness statistic changes the conclusion. The experimental protocol of the LTEE has been described elsewhere [23, 30], but we briefly summarize

the main aspects: Starting with a single ancestral strain of *E. coli*, 12 replicate populations were inoculated in 1988 and are perpetually grown in batch cultures with serial transfers, such that 1% of the population biomass is transferred to fresh growth medium each day. Samples from each replicate population are stored every 500 generations, leading to a record of evolved populations over time [30].

Previous work performed competition experiments between the ancestral population and each evolved population (every 500-2000 generations) by combining them in equal proportions and growing them over a single batch culture growth cycle, with measurements of their initial and final absolute abundances taken by colony counting [29]. This data has been prepared in a convenient format by Good et al. [31] and is available for download at `https://github.com/benjaminhgood/LTEE-metagenomic/blob/master/additional_data/Concatenated.LTEE.data.all.csv`. For a few of the 12 populations, the time series is truncated: population Ara+6 has competition measurements up to generation 4000, population Ara-2 has competition measurements up to generation 30,000, and population Ara+2 has competition measurements up to generation 32,000 [29]. Note that the evolved population tested in these competitions is not a single genotype, but a sample of multiple genotypes that were present in that evolving population at that time. From these values of absolute abundance, we compute the log fold-change (LFC) of the evolved and ancestral populations in each competition and then calculate the evolved population's relative fitness per-cycle (Eq. (S30)) and per-generation (Eq. (S31)). The original dataset has two to four replicate measurements for each evolved sample, corresponding to a repeat of the competition experiment at a different day [29]. Initially, we collect all competition experiments into a single dataset ($n = 928$ competitions) and find that relative fitness per-generation and per-cycle differ in the ranking of these competitions (Fig. S9A,B)For example, one competition is ranked 261 positions lower in relative fitness per-generation than per-cycle (where higher ranks indicate higher fitness). The scatter occurs because the biomass yield evolves downward over time [32, 33]. We can understand the ranking mismatch from the underlying LFCs in the competition experiment (Fig. S9C), that show considerable scatter, with some mutant-wild-type pairs in a positive covariation (compare Fig. S9C and Fig. 1D). The scatter occurs because the biomass yield evolves downward over time [32, 33], shifting the wild-type LFCs downward in 50:50 competitions with the evolved populations (horizontal trend across time points in Fig. S9C).

We can also construct a timeseries of relative fitness for each population and compare fitness rankings at a individual time points. For each of the 12 populations in the LTEE, we define the relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$ at time t by averaging the value $s^{\text{logit}}_{\text{gen}}$ across all competition experiments with the frozen sample from time t. Simlarly, we define a timeseries of relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ for each population. In summary, we can pool all time series into a single dataset (with the truncation described above) and test how the two statistics rank the 12 populations at any point in the experiment. We find that the mismatch between relative fitness per-cycle and per-generation is consistenly low at all time points (Fig. S10), with a few exceptions (for example, at generation 4,000 the two statistics disagree on the top six populations).

A key result from previous analysis on this dataset is that the evolving populations increase indefinitely in relative fitness, rather than leveling off at some maximum fitness value. The original analysis by Wiser et al. [29] tested the long-term trend by fitting the statistic $W = 1 + s^{\text{logit}}_{\text{gen}}$ to a hyperbolic model of the time series:

$$f(t) = 1 + \frac{at}{t+b}, \tag{S43}$$

where t is the evolutionary time point at which the evolved population is measured against its ancestor, and $f(t)$ is the fitness statistic measured from that competition (here fitted to the measured values of W). The important feature of the hyperbolic model is that it assumes that relative fitness saturates at a maximum fitness over long times ($\lim_{t\to\infty} f(t) = 1 + a$). To contrast this model, they also tested a power law

$$f(t) = (\alpha t + 1)^{\beta} \tag{S44}$$

under which fitness increases without bound over long times ($\lim_{t\to\infty} f(t) = \infty$). To repeat this analysis with the fitness statistics $s^{\text{logit}}_{\text{cycle}}$ and $s^{\text{logit}}_{\text{gen}}$ used in this article, we must adjust the models to account for the fact that W takes 1 as its neutral value (occurring at $t = 0$ by definition) while $s^{\text{logit}}_{\text{cycle}}$ and $s^{\text{logit}}_{\text{gen}}$ are zero under neutrality. Thus we use

$$f(t) = \frac{at}{t+b} \tag{S45}$$

as the hyperbolic model and

$$f(t) = (\alpha t)^{\beta} \tag{S46}$$

as the power law model.

As a control against the original analysis of Wiser et al. [29], we first perform our own fit of the time series of the relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$ to the hyperbolic (Eq. (S45)) and the power law (Eq. (S46)) models. Wiser et al. compared their two models using the Bayesian Information Criterion [29], but since the models have the same number of parameters this is mathematically equivalent to comparing the values of R^2. Figure S11A shows that in our analysis, the power law model has a higher quality of fit ($R^2 = 0.701$) than the hyperbolic model does ($R^2 = 0.682$) for the fitness statistic $s^{\text{logit}}_{\text{gen}}$, consistent with the original result by Wiser et al. for the statistic $W = 1 + s^{\text{logit}}_{\text{gen}}$ [29]. Since the equations

for the hyperbolic model only differ in the constant offset (compare Eqs. (S43) and (S45)), our fit of $s_{\text{gen}}^{\text{logit}}$ to Eq. (S45) is mathematically equivalent to the fit of W to Eq. (S43). Therefore, our fit should give the exact same results as the original publication [29] in the case of the hyperbolic model (but we couldn't fit the fitted values a, b in [29] so we were unable to check it). This is not true for the power law, because the modified power law used by Wiser et al. [29] (Eq. (S44)) includes the offset within the parentheses, rather than as an added constant outside (compare Eq. (S44) to Eq. (S46)). This means that the fit of W to the power law with an initial value of 1 (Eq. (S44), fitted values $\alpha = 0.00515, \beta = 0.0950$; matching [29]) is different from the fit of $s_{\text{gen}}^{\text{logit}}$ to the power law with an initial value of zero (Eq. (S46), fitted values $\alpha = 0.000007, \beta = 0.299891$).

We next perform a new analysis by calculating the relative fitness per-cycle $s_{\text{cycle}}^{\text{logit}}$ from the same timeseries data (12 timeseries pooled together, one for each line) and fitting this fitness statistic to the same hyperbolic (Eq. (S45)) and power law (Eq. (S46)) models. As shown in Fig. S11B, we find that the power law model outperforms the hyperbolic model for the per-cycle fitness, consistent with the model performance for the per-generation fitness. This suggests that Wiser et al.'s original conclusion about fitness increasing without bound [29] is robust to the choice of fitness statistic.

As we previously mentioned, the fitness measurements of all replicate populations are not uniform across time (Fig. S11A): there are fewer fitness measurements at late time points (generation 34,000 and higher) because three populations were eventually excluded from the fitness measurements [29]. To further corroborate our results, we thus repeat the model fits using a single time series of the evolved fitness, rather than fitting the models to a all 12 fitness time series simultaneously. We calculate the average fitness per-generation $s_{\text{gen}}^{\text{logit}}$ and per-cycle $s_{\text{cycle}}^{\text{logit}}$ in the evolution experiment as the average across all 12 populations at each time point and fit this population-averaged time series to the hyperbolic (Eq. (S45)) and power law (Eq. (S46)) models (Wiser et al [29] refer to this as a fit to the "grand mean"). Figure S11C,D shows that the power law still has a better quality of fit than the hyperbolic model does, and we thus conclude that the increasing fitness trend reported by Wiser et al. [29] is also robust to the uneven distribution of measurements over time.

We note that the quality of fit R reported in the main text of Wiser et al. [29] differs from the quality of fit we show on our plots. While both studies performed the fits of the models on all 12 populations simultaneously, Wiser et al. [29] evaluated the quality of fit by calculating R for the fitted model (blue and pink lines in Fig. S11A) against the fitness time series averaged over replicates (grey points in Fig. S11C). That is, they fit the model to the data without averaging over population but calculate quality of fit using the population-averaged data. Since we believe this was inconsistent, we have followed a more standard approach of calculating the quality of fit on the same input data used for the fit (i.e., correlating the blue and pink lines in Fig. S11A with grey points in the same plot). As a consequence, the correlations between the fitted models and data reported in the original publication (hyperbolic $R = 0.969$, power law $R = 0.986$; see [29]) are systematically higher than the values we find in our re-analysis (hyperbolic $R = 0.826$, power law: $R = 0.837$; take square root of the values in Fig. S11A). Although the long-term fitness dynamics in the case of the LTEE are robust to the choice of fitness statistic, we note that it is possible to construct scenarios of microbial growth trait evolution where relative fitness saturates in one statistic but not in the other (Fig. 3). For theory work that on the long-term trend and other models than the powerlaw or hyperbolic model tested here, see [34–36].

S9. TESTING AUC AND OTHER FITNESS POTENTIALS USING SIMULATED COMPETITION EXPERIMENTS

In this section we explain our tests of estimated fitness potentials against true relative fitness in simulated competition experiments. We first focus on the area under the growth curve (AUC), which is defined as

$$\text{AUC} = \int_0^{t_{\text{eval}}} N(t)dt, \tag{S47}$$

where $N(t)$ is a growth curve of absolute abundance (or a proxy such as optical density) and t_{eval} is a cut-off time for evaluating the area. Many previous studies [12, 15, 16, 37, 38] and growth curve analysis packages [39] have used this definition. The idea of AUC is that, unlike estimated fitness potentials that only account for individual traits of a genotype's growth (e.g., growth rate or lag time; see below), the AUC literally integrates the whole growth dynamics into a single number. For example, both fast growth rate and short lag time are manifested in greater AUC. Note that while one can attempt to use the AUC as an approximate fitness potential (as we investigate here), it is not a measure of absolute fitness (Sec. S3) since it is insufficient to predict changes in absolute abundance (i.e., the area under the growth curve does not determine the change in absolute abundance from beginning to end).

To compute the AUC for the set of single-gene deletion genotypes in our data set, we first simulate a growth curve for each genotype under the population dynamics model in Eq. (11) with the traits estimated from the original data (Methods). We use simulated growth curves, with trait values inferred from the measured growth curves, rather than the actual measured growth curves since we are comparing the AUCs to relative fitness also from simulations of competitions; we do not have actual competition data for these genotypes, so it would be an apples-to-oranges comparison if we used AUCs from the actual growth curve data. Furthermore, in empirical

growth curves, the AUC is also influenced by technical variation in the initial biomass N_0 and the initial concentration of resources R_0, so using simulations removes these effects.

Figure S14A shows the distribution of saturation times $t_{\rm sat}$ (as defined for the population dynamics model in Methods) numerically calculated for all simulated growth curves of the deletion mutants. We calculate the AUC for each growth curve using Eq. (S47) with an evaluation time of $t_{\rm eval} = 16$ hours, since that includes the stationary phase in the vast majority of our simulated growth curves (Fig. S14A). From the mutant's AUC we calculate the AUC-based estimator of the mutant's relative fitness

$$\hat{s}_{\rm AUC} = {\rm AUC}_{\rm mut} - {\rm AUC}_{\rm wt}, \tag{S48}$$

where ${\rm AUC}_{\rm wt}$ is the AUC for a simulated growth curve of the wild-type (using the median wild-type traits in our database; see Methods). We then simulate a single growth cycle of the mutant competing against the wild-type (Eq. (11) with equal initial abundance of the mutant and wild-type) to calculate the mutant's "true" (under the model assumptions) relative fitness per-cycle with the logit encoding $s^{\rm logit}_{\rm cycle}$. We simulate both the coculture and the monoculture growth dynamics with the same initial biomass $N_0 = 0.05$ OD and resource concentration $R_0 = 111$ mM [24]. Repeating this analysis for all mutant genotypes in the dataset, we calculate the Spearman rank correlation between the AUC-predicted relative fitness from monoculture $\hat{s}_{\rm AUC}$ and the true relative fitness in competition $s^{\rm logit}_{\rm cycle}$ (Fig. S13, column C).

The success of the AUC depends on the evaluation time $t_{\rm eval}$, which sets the time window from which information is captured from the growth curve by the integral (Eq. (S47)). Figure S14B–D shows the correlation between the relative fitness in coculture $s^{\rm logit}_{\rm cycle}$ and the AUC estimator $\hat{s}_{\rm AUC}$ (Eq. (S48)) for three values of $t_{\rm eval}$: the mean saturation time of all genotypes in monoculture ($t_{\rm eval} \approx 13$ hours), a significant longer value ($t_{\rm eval} \approx 24$ hours), and an intermediate value ($t_{\rm eval} = 16$ hours; used for Fig. S13). For short evaluation times ($t_{\rm eval} \approx 13$ hours), the AUC underestimates the fitness of mutants with long lag but fast growth, which leads to a nonlinear relationship between the AUC estimator and the true relative fitness (compare Fig. S14B and Fig. S14C). For long evaluation times ($t_{\rm eval} = 24$ hours), there is greater scatter between the AUC predictor and true relative fitness for the highest fitness mutants (compare spread in Fig. S14D and Fig. S14C). Intuitively, these mutants have short lag time or fast growth rate and saturate early in monoculture, so their AUC values are effectively set by the biomass yield, which has no predictive value on the competition outcome. In summary, Fig. S14B–D shows a trade-off between accurately ranking highly-deleterious mutants (which needs long $t_{\rm eval}$) and ranking highly-beneficial mutants (which need short $t_{\rm eval}$).

Besides AUC it is possible to use other features of the monoculture growth curve as approximate fitness potentials. For example, one can use the monoculture growth rates alone as fitness potentials:

$$\hat{s}_g = g^{\rm mono}_{\rm mut} - g^{\rm mono}_{\rm wt}, \tag{S49}$$

or the monoculture lag times:

$$\hat{s}_{\rm lag} = \lambda^{\rm mono}_{\rm mut} - \lambda^{\rm mono}_{\rm wt}. \tag{S50}$$

Another possible fitness potential is the absolute abundance at saturation, which in our model of population dynamics (Eq. (11)) is proportional to the biomass yield:

$$\hat{s}_Y = Y^{\rm mono}_{\rm mut} - Y^{\rm mono}_{\rm wt}. \tag{S51}$$

Finally, one can also use the difference in the monoculture log fold-changes (LFCs):

$$\hat{s}_{\rm LFC} = \log\left(\frac{N_{\rm mut}(t^{\rm mono}_{\rm sat})}{N_{\rm mut}(0)}\right) - \log\left(\frac{N_{\rm wt}(t^{\rm mono}_{\rm sat})}{N_{\rm wt}(0)}\right). \tag{S52}$$

This looks similar to the definition of relative fitness per-cycle (Eq. (S30)) but is distinct because it uses the LFCs in monoculture rather than the true LFCs in coculture (which may be different).

We test each of these fitness potentials against the relative fitness in pairwise competition, using different input datasets of trait variation and the 'GNU Parallel' command to speed up the simulation process [40]. Figure S13 rows B and C show that growth rate and lag time can act as perfect fitness potentials if that trait is the only trait with variation across mutants. This is because the relative fitness is proportional to differences in each of these traits when they are the only source of variation (see Sec. S13). Section S13 also shows that differences in biomass yield ($Y_{\rm mut} - Y_{\rm wt}$) have no effect on fitness by themselves, which is why the biomass yield of the strains in monoculture is a poor fitness potential to estimate relative fitness. This large but neutral variation in the biomass yield across mutants in our datasets means that the LFC is also a poor fitness potential.

All of these trait-based fitness potentials are outperformed by the AUC, which provides the best approximation of the mutant fitness ranking in coculture under realistic trait variation (Fig. S13, row A). More broadly, it is important to treat absolute and relative fitness, as well as fitness potentials, as distinct concepts, serving different purposes [1, 41]. As we show here in simulation (Fig. S13), and others have shown in experiments [15, 16, 42, 43], measuring fitness potentials is not enough to demonstrate that a mutant genotype will outcompete the wild-type.

S10. COARSE-GRAINING PAIRWISE RELATIVE FITNESS IN MULTI-GENOTYPE POPULATIONS

In this section, we point out the specific advantages of the logit-encoding for coarse-graining pairwise relative

fitness in bulk competition experiments. While the pairwise relative fitness is defined for any encoding m, the logit encoding endows it with some convenient mathematical properties not shared by other encodings (e.g., the log encoding). The logit encoding of the pairwise relative abundance has the property

$$\operatorname{logit} x_{ij} = \log x_i - \log x_j, \tag{S53}$$

meaning that it is antisymmetric under exchange of the indices i and j ($\operatorname{logit} x_{ij} = -\operatorname{logit} x_{ji}$) and additive across pairs of indices ($\operatorname{logit} x_{ij} = \operatorname{logit} x_{ik} + \operatorname{logit} x_{kj}$). Since the logit-encoded pairwise relative fitness is just the time derivative of the logit function (Eq. (19)), it carries equivalent properties of antisymmetry and additivity:

$$s_{ji}^{\text{logit}} = -s_{ij}^{\text{logit}}, \tag{S54}$$

$$s_{ij}^{\text{logit}} = s_{ik}^{\text{logit}} + s_{kj}^{\text{logit}}. \tag{S55}$$

We also note that the logit encoding of the pairwise relative abundance has the property:

$$\operatorname{logit}\left(\frac{ax_i}{ax_i + bx_j}\right) = \operatorname{logit}\left(\frac{x_i}{x_i + x_j}\right) + \log\left(\frac{a}{b}\right). \tag{S56}$$

Rescaling the relative abundances of either genotype thus does not change the pairwise relative fitness (since it only shifts the logit by a constant, which does not affect its derivative). This means that pairwise relative fitness is an "intensive" property of a genotype, analogous to intensive properties in statistical mechanics (such as temperature) that do not scale with system size. For example, if we split a mutant genotype into two subgroups (e.g., differentiated by a neutral marker), the pairwise relative fitness of each mutant subgroup with respect to the wild-type will be the same as the pairwise relative fitness of the mutant genotype as a whole compared to the wild-type. In contrast, the logit-encoded total relative fitness does not satisfy this property since $\operatorname{logit}(ax_i) \neq \operatorname{logit} x_i +$ constant.

When the encoding m is the logit function, the pairwise relative fitness per-cycle still satisfies the above properties (antisymmetry and additivity with respect to indices i, j, and invariance under relative abundance rescaling) since those are properties of the underlying logit encoding. This is also apparent from interpreting the per-cycle fitness statistic as an integral of the instantaneous statistic:

$$s_{ij,\text{cycle}}^{m} = \int_{t_0}^{t_{\text{sat}}} s_{ij}^{m}(t)\, dt. \tag{S57}$$

S11. THE RELATIVE FITNESS BETWEEN COARSE-GRAINED GROUPS OF GENOTYPES

In this section, we generalize the concept of pairwise relative fitness to pairs of genotype groups rather than pairs of individual genotypes. In a multi-genotype population with non-overlapping subsets of genotypes $\mathcal{A}$ and $\mathcal{B}$, define

$$x_{\mathcal{AB}}(t) = \frac{\sum_{i\in\mathcal{A}} x_i(t)}{\sum_{i\in\mathcal{A}} x_i(t) + \sum_{i\in\mathcal{B}} x_i(t)}. \tag{S58}$$

as the relative abundance of $\mathcal{A}$ genotypes compared to $\mathcal{B}$ genotypes at time t. Analogous with Eq. (19), we define the fitness of group $\mathcal{A}$ relative to group $\mathcal{B}$ as

$$s_{\mathcal{AB}}^{m} = \frac{d}{dt} m(x_{\mathcal{AB}}(t)) \tag{S59}$$

for an encoding m. Under the logit encoding, it turns out that the fitness between these two groups can be conveniently expressed as a weighted average of the pairwise fitness between the member genotypes in each group:

$$s_{\mathcal{AB}}^{\text{logit}} = \frac{1}{\sum_{i\in\mathcal{A}} x_i} \frac{1}{\sum_{j\in\mathcal{B}} x_j} \cdot \sum_{i\in\mathcal{A}} \sum_{j\in\mathcal{B}} x_i x_j s_{ij}^{\text{logit}}. \tag{S60}$$

To prove Eq. (S60), we first note that

$$\frac{x_{\mathcal{AB}}}{1 - x_{\mathcal{AB}}} = \frac{\sum_{i\in\mathcal{A}} x_i}{\sum_{j\in\mathcal{B}} x_j}. \tag{S61}$$

Thus we can rewrite the logit-encoded relative pairwise fitness between $\mathcal{A}$ and $\mathcal{B}$ as

$$\begin{aligned} s_{\mathcal{AB}}^{\text{logit}} &= \frac{d}{dt} \operatorname{logit} x_{\mathcal{AB}} \\ &= \frac{d}{dt} \log\left(\frac{\sum_{i\in\mathcal{A}} x_i}{\sum_{j\in\mathcal{B}} x_j}\right) \\ &= \frac{d}{dt} \log\left(\sum_{i\in\mathcal{A}} x_i\right) - \frac{d}{dt} \log\left(\sum_{j\in\mathcal{B}} x_j\right), \end{aligned} \tag{S62}$$

where we have invoked Eq. (S61) on the second line. We can expand each term on the right-hand side of Eq. (S62) as

$$\frac{d}{dt} \log \sum_{i\in\mathcal{A}} x_i = \frac{1}{\sum_{i\in\mathcal{A}} x_i} \sum_{i\in\mathcal{A}} \frac{dx_i}{dt}, \tag{S63}$$

and then insert Eq. (S63) into Eq. (S62) to obtain

$$\begin{aligned} s_{\mathcal{AB}}^{\text{logit}} =& \frac{1}{\sum_{i\in\mathcal{A}} x_i} \frac{1}{\sum_{j\in\mathcal{B}} x_j} \\ &\cdot \left[\sum_{j\in\mathcal{B}} x_j \sum_{i\in\mathcal{A}} \frac{dx_i}{dt} - \sum_{i\in\mathcal{A}} x_i \sum_{j\in\mathcal{B}} \frac{dx_j}{dt}\right], \end{aligned} \tag{S64}$$

where we have collected the normalization factors (sums over relative abundances in $\mathcal{A}$ and $\mathcal{B}$) as a single prefactor. We then rewrite each product of sums in Eq. (S64) as

$$\sum_{j\in\mathcal{B}} x_j \sum_{i\in\mathcal{A}} \frac{dx_i}{dt} = \sum_{j\in\mathcal{B}} \sum_{i\in\mathcal{A}} x_j x_i \frac{d}{dt} \log x_i \tag{S65}$$

to finally obtain

$$s^{\text{logit}}_{\mathcal{AB}} = \frac{1}{\sum_{i\in\mathcal{A}} x_i}\frac{1}{\sum_{j\in\mathcal{B}} x_j} \cdot \left[\sum_{j\in\mathcal{B}}\sum_{i\in\mathcal{A}} x_j x_i \left(\frac{d}{dt}\log x_i - \frac{d}{dt}\log x_j\right)\right]. \tag{S66}$$

Identifying the term in the inner parentheses as the pairwise selection coefficient s^{logit}_{ij} (Eq. (19)) then results in Eq. (S60).

Equation (S60) establishes that the relative fitness between a pair of genotype groups is a weighted sum of relative fitnesses between individual pairs of genotypes in those groups, but this holds for relative fitness defined at an instant in time (since it is based on derivatives). For relative fitness defined over a finite time interval (e.g., a growth cycle in batch culture), an analogous but approximate result holds. We first write the relative fitness over a growth cycle time interval as an integral over the instantaneous relative fitness (inserting Eq. (S60) into Eq. (S57)):

$$s^{\text{logit}}_{\mathcal{AB},\text{cycle}} = \int_0^{t_{\text{sat}}} \frac{1}{\sum_{i\in\mathcal{A}} x_i(t)}\frac{1}{\sum_{j\in\mathcal{B}} x_j(t)} \cdot \sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{B}} x_i(t)x_j(t)s^{\text{logit}}_{ij}(t)dt, \tag{S67}$$

where t_{sat} is the end time of the growth cycle (Methods, Sec. S7). The integral in Eq. (S67) is difficult to calculate as the relative abundance trajectories $x_i(t), x_j(t)$ depend on the relative fitness of the genotypes $s^{\text{logit}}_{ij}(t)$ in a non-trivial way. Instead, we make the approximation that the relative abundances do not change significantly over time of the growth cycle ($x_i(t) \approx x_i(0)$) and can thus pass the integral through the sums in Eq. (S67) to show that the per-cycle relative fitness between a pair of genotype groups is approximately also a weighted sum of per-cycle relative fitnesses:

$$s^{\text{logit}}_{\mathcal{AB},\text{cycle}} \approx \frac{1}{\sum_{i\in\mathcal{A}} x_i(0)}\frac{1}{\sum_{j\in\mathcal{B}} x_j(0)} \cdot \sum_{i\in\mathcal{A}}\sum_{j\in\mathcal{B}} x_i(0)x_j(0)s^{\text{logit}}_{ij,\text{cycle}}. \tag{S68}$$

Conceptually, assuming the relative abundances are approximately constant over the growth cycle is equivalent to assuming selection is weak; one can also show this mathematically by expressing the relative abundances $x_i(t)$ in terms of the pairwise relative fitnesses s^{logit}_{ij} in Eq. (S67) and keeping only terms to leading order in s^{logit}_{ij}.

We finally note that the total relative fitness (instantaneous Eq. (17) and per-cycle Eq. (20)) is a special case of the relative fitness between groups (Eqs. (S60) and (S68)) where $\mathcal{A}$ is the single genotype i and $\mathcal{B}$ is all other genotypes besides i:

$$\sigma^{\text{logit}}_i = \frac{1}{1-x_i}\cdot\sum_{j\neq i} x_j s^{\text{logit}}_{ij}, \tag{S69}$$

$$\sigma^{\text{logit}}_{i,\text{cycle}} \approx \frac{1}{1-x_i(0)}\cdot\sum_{j\neq i} x_j(0) s^{\text{logit}}_{ij,\text{cycle}}. \tag{S70}$$

S12. FITNESS ERROR FROM THE FRAME OF REFERENCE IN BULK COMPETITION EXPERIMENTS

Here we calculate the error that arises from measuring the total relative fitness of each mutant in a bulk competition experiment of a mutant library, rather than the pairwise relative fitness between each mutant and the wild-type. We call this difference the *error from the frame of reference*, the frame of reference being either the whole population in total fitness or the wild-type in pairwise fitness. Note that this is an error between two different fitness quantifications of the same bulk competition experiment; Sec. S14 addresses the error (arising from higher-order interactions) between fitness quantifications in bulk versus pairwise competition experiments. Here we only consider relative fitness under the logit encoding and measured per growth cycle, so we drop these labels to simplify notation.

Consider a bulk competition experiment of a wild-type and a library of large number of mutants over a single batch growth cycle of time t_{sat} (Methods). The total relative fitness of mutant i is (Eq. (20))

$$\sigma_i = \text{logit } x_i(t_{\text{sat}}) - \text{logit } x_i(0), \tag{S71}$$

while its pairwise relative fitness compared to the wild-type is (Eq. (21))

$$s_{i,\text{wt}} = \text{logit } x_{i,\text{wt}}(t_{\text{sat}}) - \text{logit } x_{i,\text{wt}}(0). \tag{S72}$$

Using the coarse-graining rules from Sec. S11 (namely Eqs. (S68) and (S70)), we can express the total relative fitness of mutant i as a weighted sum of the pairwise relative fitnesses between the mutant and the wild-type and between the mutant and the rest of the mutant library:

$$\sigma_i \approx \frac{1}{x_{\text{wt}} + x_{\text{lib}\setminus i}}\left(x_{\text{wt}}s_{i,\text{wt}} + x_{\text{lib}\setminus i}s_{i,\text{lib}\setminus i}\right), \tag{S73}$$

where the notation $\text{lib}\setminus i$ refers to the mutant library excluding the mutant i. The approximation here is due to our assumption in Eq. (S68) that selection is weak enough that the relative abundances of genotypes do not change too much over the growth cycle. Since we can rewrite the pairwise relative fitness between mutant i and the rest of the library as a difference between i and the wild-type and the rest of the library and the wild-type (using Eq. (S55))

$$s_{i,\text{lib}\setminus i} = s_{i,\text{wt}} - s_{\text{lib}\setminus i,\text{wt}}, \tag{S74}$$

we insert this into Eq. (S73) to obtain

$$\sigma_i \approx s_{i,\mathrm{wt}} - \frac{x_{\mathrm{lib}\setminus i}}{x_{\mathrm{wt}} + x_{\mathrm{lib}\setminus i}} s_{\mathrm{lib}\setminus i,\mathrm{wt}}. \tag{S75}$$

The difference between the total and pairwise relative fitness is therefore

$$\sigma_i - s_{i,\mathrm{wt}} \approx -\frac{x_{\mathrm{lib}\setminus i}}{x_{\mathrm{wt}} + x_{\mathrm{lib}\setminus i}} s_{\mathrm{lib}\setminus i,\mathrm{wt}}. \tag{S76}$$

Since mutant libraries in these experiments typically contain hundreds or thousands of mutants, the contribution of a single mutant i is small and thus we can assume that the properties of the mutant library excluding mutant i ($\mathrm{lib} \setminus i$) are approximately the same as the library as a whole (so that $x_{\mathrm{lib}\setminus i} \approx x_{\mathrm{lib}}$, $s_{\mathrm{lib}\setminus i,\mathrm{wt}} \approx s_{\mathrm{lib},\mathrm{wt}}$). This allows us to further simplify the error as

$$\sigma_i - s_{i,\mathrm{wt}} \approx -x_{\mathrm{lib}} s_{\mathrm{lib},\mathrm{wt}}. \tag{S77}$$

Equation (S77) shows why this offset between the total and pairwise relative fitnesses is approximately independent of the focal mutant i (hence the constant shift for all mutant points in main text Fig. 4C). The sign of the error from the frame of reference depends on the mean fitness of the mutant library: a mutant library that is overall deleterious relative to the wild-type ($s_{\mathrm{lib},\mathrm{wt}} < 0$) causes the total relative fitness for a mutant σ_i to overestimate the mutant's pairwise relative fitness $s_{i,\mathrm{wt}}$. Intuitively, this is because the total relative fitness is comparing the mutant to a mixed population of wild-type and other mutants, which are on average worse competitors than the wild-type, which thus makes the mutant appear to be better than if it is just compared to the wild-type alone (compare top and bottom panel in Fig. 4B). Equation (S77) also shows that the error from the frame of reference can be reduced if the mutant library is neutral relative to the wild-type ($s_{\mathrm{lib},\mathrm{wt}} = 0$) or the mutant library has small relative abundance in the culture biomass ($x_{\mathrm{lib}} \ll 1$). In bulk competition experiments with barcoded mutant libraries, these assumptions are often not met since the mutant libraries tend to be overall deleterious (as in our simulated bulk competition for the yeast single-gene knockouts, see Fig. S6) and are inoculated at a high relative abundance [44–48]. Since this makes the error in Eq. (S77) significant, we instead recommend including barcoded wild-type strains as references in the bulk competition, so that pairwise fitness can be quantified relative to that (Eq. (S72)) rather than using the total relative fitness. By using a mix of barcoded wild-type cells and non-barcoded wild-type cells it is further possible to optimize this protocol and save on sequencing investment [49].

Finally, we want to point out a difference between the best practice we recommend here (Discussion) and a wide-spread practice in estimating fitness estimates in bulk competition experiments. In practice, transposon-seq experiments that grow the mutant library by itself start with an estimate of total relative fitness, and then subtract the median total relative fitness of the knockouts [45–47, 50] or a mean total relative fitness [44, 51, 52]. However, these corrections are not explicitly founded in the choice of a reference group (like a set of neutral genotypes or a wild-type), making the correction appear ad-hoc [45, 47, 50], or the reference is a strain that is not part of the culture, like in the fitness estimates for barcoded lineages that are evaluated against the initial ancestor without that ancestor actually being present [44, 51, 52]. To make things more confusing, even those studies that do include a wild-type then describe their method as an estimate of total relative fitness of the mutant under the log encoding, subtracted by the total relative fitness of the wild-type under the log-encoding [49, 53, 54]. The result is a pairwise relative fitness under the logit-encoding (Eq. (23)) but presenting it this way obscures that choice of encoding and the relationship to the classic, logit-based selection coefficient used in pairwise competition experiments (Eq. (6)). We hope that our framework can provide more clarity: The choice of the reference group happens at the level of relative abundance, by calculating a pairwise relative abundance (Eq. (18) or Eq. (S58)), and this removes the need for any correction on the fitness values themselves.

S13. PAIRWISE RELATIVE FITNESS USING AN EXPLICIT MODEL OF POPULATION DYNAMICS

In the model of population dynamics (Methods, Eq. (11)), we can calculate the pairwise relative fitness of genotypes based on the approximation of the saturation time (Sec. S7; [27, 28]). The pairwise relative fitness of genotype i relative to genotype j (per-cycle and under the logit encoding) is

$$s^{\mathrm{logit}}_{ij,\mathrm{cycle}} \approx s_{ij,\mathrm{lag}} + s_{ij,\mathrm{growth}} + s_{ij,\mathrm{coupling}}, \tag{S78a}$$

where

$$s_{ij,\mathrm{lag}} = -\frac{\bar{\tau}}{\tau_i \tau_j} \Delta\lambda_{ij}, \tag{S78b}$$

$$s_{ij,\mathrm{growth}} = -\frac{\bar{\tau}}{\tau_i \tau_j} \Delta\tau_{ij} \mathrm{LFC}, \tag{S78c}$$

$$s_{ij,\mathrm{coupling}} = -\frac{\bar{\tau}\bar{Y}}{\tau_i \tau_j} \times \sum_k \frac{x_k}{\tau_k Y_k} \left(\Delta\tau_{ik}\Delta\lambda_{kj} - \Delta\lambda_{ik}\Delta\tau_{kj}\right). \tag{S78d}$$

The e-fold growth time for genotype j is $\tau_j = 1/g_j$, and the terms $\Delta\lambda_{ij} = \lambda_i - \lambda_j$ and $\Delta\tau_{ij} = \tau_i - \tau_j$ are the differences between the two genotypes lag times and growth times. Since the terms in Eq.(S78d) depend on the covariation between growth and lag, we interpret these terms as couplings between the growth and lag phases; they are zero if only two genotypes are present.

S14. FITNESS ERROR FROM HIGHER-ORDER INTERACTIONS BETWEEN PAIRWISE AND BULK COMPETITION EXPERIMENTS

In this section, we calculate the error in relative fitness of a mutant arising from higher-order interactions in bulk competition experiments with large mutant libraries, compared to the "true" relative fitness in pairwise competitions between just the focal mutant and wild-type alone. Here we only consider relative fitness under the logit encoding and measured per growth cycle, so we drop these labels to simplify notation. Let the pairwise relative fitness of mutant i compared to the wild-type (Eq. (21)) be

$$s^{\text{pair}}_{i,\text{wt}} = \text{logit } x^{\text{pair}}_{i,\text{wt}}(t^{\text{pair}}_{\text{sat}}) - \text{logit } x_{i,\text{wt}}(0) \tag{S79}$$

in the pairwise competition with the wild-type alone, and

$$s^{\text{bulk}}_{i,\text{wt}} = \text{logit } x^{\text{bulk}}_{i,\text{wt}}(t^{\text{bulk}}_{\text{sat}}) - \text{logit } x_{i,\text{wt}}(0) \tag{S80}$$

in the bulk competition with all other mutants in the library. The superscripts pair and bulk indicate that the dynamics of $x_{i,\text{wt}}(t)$ and the saturation time t_{sat} may be different in the two competitions. The difference between these two measurements of relative fitness is

$$\begin{aligned}\Delta s_{i,\text{wt}} &= s^{\text{bulk}}_{i,\text{wt}} - s^{\text{pair}}_{i,\text{wt}} \\ &= s^{\text{pair}}_{i,\text{wt}}\left(\frac{s^{\text{bulk}}_{i,\text{wt}}}{s^{\text{pair}}_{i,\text{wt}}} - 1\right).\end{aligned} \tag{S81}$$

Since the difference between the bulk and pairwise competitions is the presence of the other mutants in the library, we interpret this difference as the *fitness error from higher-order interactions* among the mutants.

We now calculate how this error depends on the underlying growth traits of the genotypes using the population dynamics model (Methods, Eq. (11)); and the explicit expression for relative fitness from Sec. S13. Based on the approximate pairwise relative fitness in this model (Eq. (S78)), the relative fitness in the pairwise competition is the sum of two terms

$$s^{\text{pair}}_{i,\text{wt}} = s^{\text{pair}}_{i,\text{wt,lag}} + s^{\text{pair}}_{i,\text{wt,growth}}, \tag{S82}$$

while the relative fitness in the bulk competition is the sum of three terms

$$s^{\text{bulk}}_{i} = s^{\text{bulk}}_{i,\text{lag}} + s^{\text{bulk}}_{i,\text{growth}} + s^{\text{bulk}}_{i,\text{coupling}}, \tag{S83}$$

where the third term represents the coupling between growth and lag phases present only in populations with more than two genotypes (Eq. (S78d)). We define the higher-order effects on the selection for lag time as

$$\Delta s_{i,\text{wt,lag}} = s^{\text{bulk}}_{i,\text{wt,lag}} - s^{\text{pair}}_{i,\text{wt,lag}}. \tag{S84}$$

Using Eq. (S78b), we can express this in terms of the underlying traits as

$$\Delta s_{i,\text{wt,lag}} = s^{\text{pair}}_{i,\text{wt,lag}} \cdot \frac{\Delta\bar{\tau}_i}{\bar{\tau}^{\text{pair},i}} \tag{S85}$$

where $\Delta\bar{\tau}_i = \bar{\tau}^{\text{bulk}} - \bar{\tau}^{\text{pair},i}$ is the difference in effective e-fold growth times (Eq. (S40)) in the bulk competition and in the pairwise competition of mutant i and the wild-type.

We similarly define the higher-order effects in the selection on growth rate as

$$\Delta s_{i,\text{wt,growth}} = s^{\text{bulk}}_{i,\text{wt,growth}} - s^{\text{pair}}_{i,\text{wt,growth}}, \tag{S86}$$

and calculate $s^{\text{bulk}}_{i,\text{wt,growth}}$ and $s^{\text{pair}}_{i,\text{wt,growth}}$ from Eq. (S78c) to get the expression

$$\begin{aligned}&\Delta s_{i,\text{wt,growth}} = \\ &\quad -\frac{\Delta\tau_{i,\text{wt}}}{\tau_{\text{wt}}\tau_i}\left(\bar{\tau}^{\text{bulk}}\text{LFC}^{\text{bulk}} - \bar{\tau}^{\text{pair},i}\text{LFC}^{\text{pair},i}\right),\end{aligned} \tag{S87}$$

where LFC^{bulk} is the log fold-change (Eq. (S42)) of the total biomass in the bulk competition and $\text{LFC}^{\text{pair},i}$ is the log fold-change of the total biomass in the pairwise competition between mutant i and the wild-type. We note that total biomass growth in the bulk competition LFC^{bulk} depends on the mutant library through the effective biomass yield (Eq. (S41)), but this dependence is weak because the yield enters only logarithmically into Eq. (S42). We thus assume that the bulk and pairwise competition have equal biomass growth ($\text{LFC}^{\text{bulk}} \approx \text{LFC}^{\text{pair},i}$) for all mutants i such that the higher-order effect on the growth rate selection (Eq. (S87)) is proportional to the increase in the mean doubling time

$$\Delta s_{i,\text{wt,growth}} \approx s^{\text{pair}}_{i,\text{wt,growth}}\frac{\Delta\bar{\tau}_i}{\bar{\tau}^{\text{pair},i}}. \tag{S88}$$

Since this has the same form as the higher-order effect on lag time selection in Eq. (S85), we can combine Eq. (S88) and Eq. (S85)) into

$$\Delta s_{i,\text{wt,lag}} + \Delta s_{i,\text{wt,growth}} \approx s^{\text{pair}}_{i,\text{wt}}\frac{\Delta\bar{\tau}_i}{\bar{\tau}^{\text{pair},i}}. \tag{S89}$$

Because the effective growth time in the pairwise competition $\bar{\tau}^{\text{pair},i}$ is approximately just the effective growth time of the wild-type alone τ_{wt} (assuming the mutant is competed against it at low initial relative abundance), this means that

$$\frac{\Delta s_{i,\text{wt,lag}} + \Delta s_{i,\text{wt,growth}}}{s^{\text{pair}}_{i,\text{wt}}} \approx \left(\frac{\bar{\tau}^{\text{bulk}}}{\tau_{\text{wt}}} - 1\right). \tag{S90}$$

That is, the relative error of lag and growth selection from higher-order interactions is approximately independent of the individual mutant, but rather depends on properties of the wild-type (τ_{wt}) and the whole mutant library (through $\bar{\tau}^{\text{bulk}}$). This is why we observe an approximately constant slope across all mutants in Fig. S16 (dark orange points) when comparing the error against the pairwise relative fitness. The slight deviation from the constant slope is due to our approximation that the LFCs are the same between the bulk and pairwise competitions; this is a good approximation for the yeast deletion

library but not exactly true, and hence causes a slightly different scaling between the growth Δs_{growth} and lag Δs_{lag} error terms compared to Eq. (S90).

Finally, we note that the growth-lag coupling terms $s^{\text{bulk}}_{i,\text{wt,coupling}}$ do not have a simple scaling with pairwise relative fitness since they depend quadratically on trait differences; this is shown as the bright orange points in Fig. S16. In the main text, we therefore refer to $\Delta s_{i,\text{wt,lag}} + \Delta s_{i,\text{wt,growth}}$ as the fitness-dependent error and the $s^{\text{bulk}}_{i,\text{wt,coupling}}$ as the fitness-independent error from higher-order interactions.

S15. CHOOSING THE MUTANT LIBRARY ABUNDANCE IN BULK COMPETITION EXPERIMENTS

Section S14 showed that measuring relative fitness of a mutant in a bulk competition (with a library of other mutants also present) entails an error due to higher-order interactions among the mutants, compared to measuring relative fitness in a pairwise competition consisting of just the mutant and wild-type. Here we show how this error depends on the relative abundance of the mutant library in the bulk competition, so that we can estimate the range of library abundances that keep the error below a desired threshold.

Calculating the relative error on fitness. Let the absolute error in relative fitness from higher-order interactions for a mutant i be $\Delta s_{i,\text{wt}}$ (Eq. (S81)). Let the relative fitness of this mutant in a pairwise competition be $s^{\text{pair}}_{i,\text{wt}}$ (Eq. (S79)). Since the error $\Delta s_{i,\text{wt}}$ depends on the relative abundance x_{lib} of the whole mutant library, our goal is to determine what range of x_{lib} keeps the error in relative fitness (compared to the "true" relative fitness in the pairwise competition) below a chosen threshold ϵ:

$$\left|\frac{\Delta s_{i,\text{wt}}}{s^{\text{pair}}_{i,\text{wt}}}\right| < \epsilon. \tag{S91}$$

In Sec. S14 we calculated the dependence of $\Delta s_{i,\text{wt}}$ on the underlying growth traits. However, since here we are mainly concerned with the dependence on the library abundance x_{lib}, we present an alternative calculation that better captures that dependence.

We start by pointing out that we can express the relative fitness in pairwise (Eq. (S79)) and bulk competitions (Eq. (S80)) in terms of the saturation times for these competitions, using the explicit solution to the population dynamics model in Eq. (14):

$$s^{\text{pair}}_{i,\text{wt}} = g_i(t^{\text{pair},i}_{\text{sat}} - \lambda_i) - g_{\text{wt}}(t^{\text{pair},i}_{\text{sat}} - \lambda_{\text{wt}}), \tag{S92}$$

$$s^{\text{bulk}}_{i,\text{wt}} = g_i(t^{\text{bulk}}_{\text{sat}} - \lambda_i) - g_{\text{wt}}(t^{\text{bulk}}_{\text{sat}} - \lambda_{\text{wt}}). \tag{S93}$$

We can thus express the error from higher-order interactions (Eq. (S81)) as

$$\Delta s_{i,\text{wt}} = \Delta g_{i,\text{wt}}(t^{\text{bulk}}_{\text{sat}} - t^{\text{pair},i}_{\text{sat}}), \tag{S94}$$

where $\Delta g_{i,\text{wt}} = g_i - g_{\text{wt}}$. Equation (S94) shows that the mutant library affects the fitness of individual mutants $s^{\text{bulk}}_{i\text{wt}}$ by changing the saturation time. Mathematically, this is equivalent to the results of Sec. S14, which showed how the difference in effective e-fold growth times $\bar{\tau}$ between bulk and pairwise competitions primarily mediated the higher-order effects, but expressed in terms of the saturation time t_{sat} (which is not identical to $\bar{\tau}$ but related through Eq. (S38)). The error in Eq. (S94) is also proportional to the growth rate advantage of the mutant compared to the wild-type. In particular, mutant genotypes that only differ in lag time are not affected by the mutant library since their advantage is accrued once at the beginning of the growth cycle and therefore does not scale with the total time of competition.

According to Eq. (S38), the saturation time t_{sat} depends on the effective lag time $\bar{\lambda}$ (Eq. (S39)), effective e-fold growth time $\bar{\tau}$ (Eq. (S40)), and the log fold-change LFC (Eq. (S42)) for the competition. To simplify the calculation, we introduce a few assumptions. We assume all mutants have the same yields Y_i such that the LFCs in the bulk and pairwise competitions are identical ($\text{LFC}^{\text{bulk}} = \text{LFC}^{\text{pair}}$). We also assume that the relative abundance of the mutant in the pairwise competition is small enough that the saturation time in that case is set entirely by the wild-type traits:

$$t^{\text{pair},i}_{\text{sat}} \approx \lambda_{\text{wt}} + \tau_{\text{wt}}\text{LFC}. \tag{S95}$$

Note that this means that the difference in saturation times is independent of the specific mutant i; the only dependence of the mutant i on the overall error from higher-order interactions is through the difference in growth rates in Eq. (S94).

We can thus write the fitness error as

$$\Delta s_{i,\text{wt}} \approx \Delta g_{i,\text{wt}}\left[(\bar{\lambda}^{\text{bulk}} - \lambda_{\text{wt}}) + (\bar{\tau}^{\text{bulk}} - \tau_{\text{wt}})\text{LFC}\right]. \tag{S96}$$

The dependence on the mutant library relative abundance x_{lib} is contained within the effective traits $\bar{\lambda}^{\text{bulk}}$ and $\bar{\tau}^{\text{bulk}}$ of the bulk competition. Using the definitions in Eqs. (S39) and (S40), we can show that

$$\bar{\lambda}^{\text{bulk}} = \frac{(1-x_{\text{lib}})\lambda_{\text{wt}}g_{\text{wt}} + x_{\text{lib}}\lambda_{\text{lib}}g_{\text{lib}}}{x_{\text{lib}}g_{\text{lib}} + (1-x_{\text{lib}})g_{\text{wt}}}, \tag{S97}$$

$$\bar{\tau}^{\text{bulk}} = \frac{1}{x_{\text{lib}}g_{\text{lib}} + (1-x_{\text{lib}})g_{\text{wt}}}. \tag{S98}$$

Inserting Eqs. (S97) and (S98) into Eq. (S96), we obtain

$$\Delta s_{i,\text{wt}} = \Delta g_{i,\text{wt}}\frac{x_{\text{lib}}g_{\text{lib}}}{x_{\text{lib}}g_{\text{lib}} + (1-x_{\text{lib}})g_{\text{wt}}} \cdot \left[\Delta\lambda_{\text{lib,wt}} + \text{LFC}\Delta\tau_{\text{lib,wt}}\right], \tag{S99}$$

where $\Delta\lambda_{\text{lib,wt}} = \lambda_{\text{lib}} - \lambda_{\text{wt}}$ is the difference in lag times, and $\Delta\tau_{\text{lib,wt}} = \tau_{\text{lib}} - \tau_{\text{wt}}$ the difference in e-fold growth times (reciprocal growth rates), between the library (defined according to Eq. (S39)) and the wild-type. All the

dependence on the library relative abundance x_{lib} is now contained in the fraction outside the square brackets in Eq. (S99).

Since the relative fitness of the mutant library as a whole compared to the wild-type in bulk competition is (using Eq. (S78))

$$s^{\text{bulk}}_{\text{lib,wt}} = -\frac{\bar{\tau}^{\text{bulk}}}{\tau_{\text{lib}}\tau_{\text{wt}}}\Delta\lambda_{\text{lib,wt}} - \frac{\bar{\tau}^{\text{bulk}}}{\tau_{\text{lib}}\tau_{\text{wt}}}\Delta\tau_{\text{lib,wt}}\text{LFC}^{\text{bulk}}, \quad \text{(S100)}$$

we can rewrite Eq. (S99) as

$$\Delta s_{i,\text{wt}} = -x_{\text{lib}} s^{\text{bulk}}_{\text{lib,wt}} \frac{\Delta g_{i,\text{wt}}}{g_{\text{wt}}}, \quad \text{(S101)}$$

This shows that a mutant library that is neutral relative to the wild-type in the bulk competition ($s^{\text{bulk}}_{\text{lib,wt}} = 0$) removes the error from higher-order interactions, even when that neutrality is based on a tradeoff between growth rates and lag times ($\Delta\lambda_{\text{lib,wt}} = -\Delta\tau_{\text{lib,wt}}\text{LFC}^{\text{bulk}}$). We note that the error of higher-order interactions (Eq. (S101)) appears similar to the error from the frame of reference (Eq. (S77), but the former includes the relative mutant growth rate $\Delta g_{i,\text{wt}}/g_{\text{wt}}$ as an additional pre-factor.

Calculating the upper bound on relative abundance of the mutant library. To calculate the maximum relative library abundance that keeps the relative error below the threshold ϵ, we input the absolute error in Eq. (S99) into the relative error bound of Eq. (S91) and rearrange to isolate the dependence on the library abundance x_{lib}:

$$\frac{x_{\text{lib}} g_{\text{lib}}}{x_{\text{lib}} g_{\text{lib}} + (1 - x_{\text{lib}}) g_{\text{wt}}} < \epsilon A_i \quad \text{(S102)}$$

where

$$A_i = \left| \frac{s^{\text{pair}}_{i,\text{wt}}}{\Delta g_{i,\text{wt}}(\Delta\lambda_{\text{lib,wt}} + \text{LFC}\Delta\tau_{\text{lib,wt}})} \right| \quad \text{(S103)}$$

depends on the focal mutant i but does not depend on the library abundance x_{lib}. The left-hand side of Eq. (S102) varies between zero and one as a function of x_{lib}. This means that if $\epsilon A_i > 1$, any value of x_{lib} will satisfy Eq. (S102), meaning that the relative error on mutant i from higher-order interactions will always be less than ϵ. For example, this holds when the mutant i has the same growth rate as the wild-type ($\Delta g_{i,\text{wt}} = 0$, which causes $A_i \to \infty$), even if it varies in lag time and/or other mutants have variation in growth rates.

We thus next consider the case where $\epsilon A_i < 1$. We multiply both sides of Eq. (S102) by the denominator on the left-hand side (always positive) to obtain

$$x_{\text{lib}} g_{\text{lib}} < \epsilon A_i (x_{\text{lib}} g_{\text{lib}} + (1 - x_{\text{lib}}) g_{\text{wt}}), \quad \text{(S104)}$$

and then collect all the terms involving x_{lib} on the left hand side:

$$x_{\text{lib}}((1 - \epsilon A_i) g_{\text{lib}} + \epsilon A_i g_{\text{wt}}) < \epsilon A_i g_{\text{wt}}. \quad \text{(S105)}$$

Since $\epsilon A_i < 1$, the factor multiplying x_{lib} on the left-hand side of Eq. (S105) must be positive and thus we can divide both sides to obtain an upper bound on the library abundance x_{lib} such that the relative error on fitness is less than ϵ:

$$x_{\text{lib}} < \frac{\epsilon A_i g_{\text{wt}}}{(1 - \epsilon A_i) g_{\text{lib}} + \epsilon A_i g_{\text{wt}}}. \quad \text{(S106)}$$

Since ϵ will typically be small, we can simplify the right-hand side of Eq. (S106) by approximating it to first-order in ϵ:

$$x_{\text{lib}} \lessapprox \epsilon A_i \frac{g_{\text{wt}}}{g_{\text{lib}}}. \quad \text{(S107)}$$

This approximation holds as long as $\epsilon A_i (g_{\text{lib}} - g_{\text{wt}})/g_{\text{lib}} \ll 1$.

The maximum mutant library abundance determined by Eq. (S106) or (S107) is specific to a single mutant genotype i, meaning that the relative fitness errors for other mutants may exceed ϵ even if the inequality for mutant i is satisfied. To keep the relative fitness errors for all mutants less than ϵ, we need to choose a library abundance x^*_{lib} such that

$$\max_{\text{muts } j} \left| \frac{\Delta s_{j,\text{wt}}}{s^{\text{pair}}_{j,\text{wt}}} \right| < \epsilon. \quad \text{(S108)}$$

This means that Eq. (S107) must be satisfied for all mutants j:

$$x^*_{\text{lib}} = \min_{\text{muts } j} \epsilon A_j \frac{g_{\text{wt}}}{g_{\text{lib}}}. \quad \text{(S109)}$$

Thus we must determine the minimum value of A_j over all mutants j. Using the definition in Eq. (S103), A_j is minimized for the mutant with minimum vale of

$$\left| \frac{s^{\text{pair}}_{j,\text{wt}}}{\Delta g_{j,\text{wt}}} \right| = \left| \frac{1}{g_{\text{wt}}}\text{LFC} - g_j \frac{\Delta\lambda_{j,\text{wt}}}{\Delta g_{j,\text{wt}}} \right|, \quad \text{(S110)}$$

where we have calculated the pairwise relative fitness in Eq. (S92) using $t^{\text{pair},i}_{\text{sat}}$ from Eq. (S95). Mutants that minimize Eq. (S110) are those with tradeoffs between lag times and growth rates such that their overall pairwise relative fitness with respect to the wild-type is zero. This makes sense, since these mutants will have a relative fitness close to zero in the pairwise competition ($|s^{\text{pair}}_{j,\text{wt}}| \ll 1$), but nonzero fitness relative to the wild-type in the bulk competition as the relative selection on the lag time and growth rate shifts (e.g., because $\bar{\tau}^{\text{bulk}} > \bar{\tau}^{\text{pair}}$ in Eq. (S78)).

Special case of variation in growth rates only. To determine an even simpler estimate on the maximum mutant library abundance, we consider the special case where genotypes vary only in growth rates and not lag times. Using the definition in Eq. (S103), $A_i = g_{\text{lib}}/(g_{\text{wt}} - g_{\text{lib}})$ for all mutants i, and thus there

is a single bound on library abundance for all mutants (using Eq. (S107)):

$$x_{\text{lib}} < \epsilon \left| \frac{g_{\text{wt}}}{g_{\text{wt}} - g_{\text{lib}}} \right|. \tag{S111}$$

This is the same as Eq. (10) in the main text.

We test Eq. (10) using our simulated competition data for the yeast single-gene deletion library (Methods) with variation in all three traits: lag time, growth rate, and biomass yield (Fig. 2B-C). We compute the growth rates of the mutant library ($g_{\text{lib}} = 0.389\,\text{h}^{-1}$ using Eq. (S40)) and the wild-type ($g_{\text{wt}} = 0.406\,\text{h}^{-1}$). Using a relative error threshold of $\epsilon = 0.01$, the maximum mutant library relative abundance according to Eq. (10) is $x_{\text{lib}} = 24.6\%$. Figure 4D shows that this mutant library abundance indeed is able to keep the relative error below the threshold for mutants with high relative fitness, but, as expected, this mutant library abundance fails for mutants close to neutrality.

We furthermore compute the maximum library abundance with Eq. (S109), based on the precise trait variation in our dataset. The mutant library has an effective lag time $\lambda_{\text{lib}} = 1.95\,\text{h}$ (compared to wild-type lag time $\lambda_{\text{wt}} = 1.92\,\text{h}$) and an effective e-fold growth time $\tau_{\text{lib}} = 2.57\,\text{h}$ (compared to the wild-type $\tau_{\text{wt}} = 2.46\,\text{h}$), leading to an overall negative relative fitness $s^{\text{bulk}}_{\text{lib,wt}} = -0.198$ based on a LFC value of $\text{LFC}^{\text{bulk}} = \log 100$ (a typical fold-change in our dataset) and the limit of low relative abundance of the library. We insert these trait values into Eq. (S109) and obtain a maximum mutant library abundance $x^*_{\text{lib}} = 0.02\%$. Figure 4D shows that this much smaller mutant library abundance is able to keep the relative error below the desired threshold for all mutant genotypes. Note, however, this is still not fully exact since we have ignored the underlying variation in biomass yield in our trait dataset.

[1] Daniel Jefferson Smith, Guilhem Doulcier, Pierrick Bourrat, et al. Why there are so many definitions of fitness in models. *EcoEvoRxiv*, 10.32942/X2V61T, 2024.

[2] Susan K. Mills and John H. Beatty. The Propensity Interpretation of Fitness. *Philosophy of Science*, 46(2):263–286, 1979.

[3] Peter Takacs and Pierrick Bourrat. The arithmetic mean of what? A Cautionary Tale about the Use of the Geometric Mean as a Measure of Fitness. *Biology & Philosophy*, 37(2):12, 2022.

[4] Brian Charlesworth. Book review: Two new books on Life-History Evolution. *The Quarterly Review of Biology*, 68(2):262–264, 1993.

[5] Brian Charlesworth. *Evolution in Age-Structured Populations*. Cambridge University Press, 1994.

[6] Helen K. Alexander, Guillaume Martin, Oliver Y. Martin, and Sebastian Bonhoeffer. Evolutionary rescue: linking theory for conservation and medicine. *Evolutionary Applications*, 7(10):1161–1179, 2014.

[7] Motoo Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47(6):713–719, 1962.

[8] Z Patwa and L.M Wahl. The fixation probability of beneficial mutations. *Journal of The Royal Society Interface*, 5(28):1279–1289, 2008.

[9] Thomas F. Hansen. On the definition and measurement of fitness in finite populations. *Journal of Theoretical Biology*, 419:36–43, 2017.

[10] Michael Doebeli, Yaroslav Ispolatov, and Burt Simon. Towards a mechanistic foundation of evolutionary theory. *eLife*, 6:e23804, 2017.

[11] Karl Drlica. The mutant selection window and antimicrobial resistance. *Journal of Antimicrobial Chemotherapy*, 52(1):11–17, 2003.

[12] Kathleen Sprouffske, José Aguilar-Rodríguez, Paul Sniegowski, and Andreas Wagner. High mutation rates limit evolutionary adaptation in Escherichia coli. *PLOS Genetics*, 14(4):e1007324, 2018.

[13] Fernanda Pinheiro, Omar Warsi, Dan I. Andersson, and Michael Lässig. Metabolic fitness landscapes predict the evolution of antibiotic resistance. *Nature Ecology & Evolution*, 5(5):677–687, 2021.

[14] Juan Diaz-Colunga, Alvaro Sanchez, and C. Brandon Ogbunugafor. Environmental modulation of global epistasis in a drug resistance fitness landscape. *Nature Communications*, 14(1):8055, 2023.

[15] Sarah B. Worthan, Robert D. P. McCarthy, and Megan G. Behringer. Case Studies in the Assessment of Microbial Fitness: Seemingly Subtle Changes Can Have Major Effects on Phenotypic Outcomes. *Journal of Molecular Evolution*, 91:311—-324, 2023.

[16] Désirée A Schmitz, Tobias Wechsler, Ingrid Mignot, and Rolf Kümmerli. Predicting bacterial interaction outcomes from monoculture growth and supernatant assays. *ISME Communications*, 4(1):ycae045, 2024.

[17] J. Arjan G.M. De Visser and Joachim Krug. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, 15(7):480–490, 2014.

[18] Claudia Bank. Epistasis and Adaptation on Fitness Landscapes. *Annual Review of Ecology, Evolution, and Systematics*, 53(1):457–479, 2022.

[19] Thomas P. Smith, Tom Clegg, Emma Ransome, et al. High-throughput characterization of bacterial responses to complex mixtures of chemical pollutants. *Nature Microbiology*, 9(4):938–948, 2024.

[20] Chevin Luis-Miguel. On measuring selection in experimental evolution. *Biology Letters*, 7(2):210–213, 2011.

[21] James P. Grover. *Resource Competition*. Springer US, Boston, MA, 1997.

[22] Joachim Krug. Population Genetics and Evolution. *arXiv:1803.08474 [q-bio]*, 2018.

[23] Richard E. Lenski, Michael R. Rose, Suzanne C. Simpson, and Scott C. Tadler. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist*, 138(6):1315–1341, 1991.

[24] Jonas Warringer, Elke Ericson, Luciano Fernandez, et al. High-resolution yeast phenomics resolves different physiological features in the saline response. *PNAS*,

100(26):15724–15729, 2003.

[25] Peter S. Swain, Keiran Stevenson, Allen Leary, et al. Inferring time derivatives including cell growth rates using Gaussian processes. *Nature Communications*, 7(1):1–8, 2016.

[26] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020.

[27] Michael Manhart, Bharat V. Adkar, and Eugene I. Shakhnovich. Trade-offs between microbial growth phases lead to frequency-dependent and non-transitive selection. *Proceedings of the Royal Society B: Biological Sciences*, 285(1872):20172459, 2018.

[28] Michael Manhart and Eugene I. Shakhnovich. Growth tradeoffs produce complex microbial communities on a single limiting resource. *Nature Communications*, 9(1):1–9, 2018.

[29] M. J. Wiser, N. Ribeck, and R. E. Lenski. Long-Term Dynamics of Adaptation in Asexual Populations. *Science*, 342(6164):1364–1367, 2013.

[30] Jeffrey E. Barrick, Zachary D. Blount, Devin M. Lake, et al. Daily Transfers, Archiving Populations, and Measuring Fitness in the Long-Term Evolution Experiment with Escherichia coli. *JoVE (Journal of Visualized Experiments)*, (198):e65342, 2023.

[31] Benjamin H. Good, Michael J. McDonald, Jeffrey E. Barrick, et al. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678):45–50, 2017.

[32] Farida Vasi, Michael Travisano, and Richard E. Lenski. Long-Term Experimental Evolution in Escherichia coli. II. Changes in Life-History Traits During Adaptation to a Seasonal Environment. *The American Naturalist*, 144(3):432–456, 1994.

[33] Nkrumah A. Grant, Ali Abdel Magid, Joshua Franklin, et al. Changes in Cell Size and Shape during 50,000 Generations of Experimental Evolution with Escherichia coli. *Journal of Bacteriology*, 203(10), 2021.

[34] Yipei Guo, Marija Vucelja, and Ariel Amir. Stochastic tunneling across fitness valleys can give rise to a logarithmic long-term fitness trajectory. *Science Advances*, 5(7):eaav3842, 2019.

[35] Yipei Guo and Ariel Amir. The effect of weak clonal interference on average fitness trajectories in the presence of macroscopic epistasis. *Genetics*, 220(4):iyac028, 2022.

[36] Nicholas M Boffi, Yipei Guo, Chris H Rycroft, and Ariel Amir. How microscopic epistasis and clonal interference shape the fitness trajectory in a spin glass model of microbial long-term evolution. *eLife*, 12:RP87895, 2024.

[37] Philippe Piccardi, Björn Vessman, and Sara Mitri. Toxicity drives facilitation between 4 bacterial species. *Proceedings of the National Academy of Sciences*, 116(32):15979–15984, 2019.

[38] Andrea R. Dos Santos, Rita Di Martino, Samuele E. A. Testa, and Sara Mitri. Classifying Interactions in a Synthetic Bacterial Community Is Hindered by Inhibitory Growth Medium. *mSystems*, 7(5):e00239–22, 2022.

[39] Kathleen Sprouffske and Andreas Wagner. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, 17(1):172, 2016.

[40] Ole Tange. GNU Parallel 20231122 ('Grindavík'), 2023.

[41] Isabel Gordo, Lilia Perfeito, and Ana Sousa. Fitness Effects of Mutations in Bacteria. *Journal of Molecular Microbiology and Biotechnology*, 21(1-2):20–35, 2012.

[42] Jeniffer Concepción-Acevedo, Howard N. Weiss, Waqas Nasir Chaudhry, and Bruce R. Levin. Malthusian Parameters as Estimators of the Fitness of Microbes: A Cautionary Tale about the Low Side of High Throughput. *PLOS ONE*, 10(6):e0126915, 2015.

[43] Paulo Durão, Massimo Amicone, Lília Perfeito, and Isabel Gordo. Competition dynamics in long-term propagations of Schizosaccharomyces pombe strain communities. *Ecology and Evolution*, 11(21):15085–15097, 2021.

[44] Sasha F. Levy, Jamie R. Blundell, Sandeep Venkataram, et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181–186, 2015.

[45] Kelly M. Wetmore, Morgan N. Price, Robert J. Waters, et al. Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *mBio*, 6(3):e00306–15, 2015.

[46] Morgan N. Price, Kelly M. Wetmore, R. Jordan Waters, et al. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature*, 557(7706):503–509, 2018.

[47] Manon Morin, Emily C Pierce, and Rachel J Dutton. Changes in the genetic requirements for microbial interactions with increasing community complexity. *eLife*, 7:e37072, 2018.

[48] Emily C. Pierce, Manon Morin, Jessica C. Little, et al. Bacterial–fungal interactions revealed by genome-wide analysis of bacterial mutant fitness. *Nat Microbiol*, 6(1):87–102, 2021.

[49] Kara Schmidlin, Sam Apodaca, Daphne Newell, et al. Distinguishing mutants that resist drugs via different mechanisms by examining fitness tradeoffs. *eLife*, 13:RP94144, 2024.

[50] Jonathan N V Martinson, Jeremy M Chacón, Brian A Smith, et al. Mutualism reduces the severity of gene disruptions in predictable ways across microbial communities. *The ISME Journal*, 17(12):2270–2278, 2023.

[51] Milo S. Johnson, Alena Martsul, Sergey Kryazhimskiy, and Michael M. Desai. Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science*, 366(6464):490–493, 2019.

[52] Alex N. Nguyen Ba, Ivana Cvijović, José I. Rojas Echenique, et al. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature*, 575(7783):494–499, 2019.

[53] Christopher W Bakerlee, Angela M Phillips, Alex N Nguyen Ba, and Michael M Desai. Dynamics and variability in the pleiotropic effects of adaptation in laboratory budding yeast populations. *eLife*, 10:e70918, 2021.

[54] Alejandro Couce, Anurag Limdi, Melanie Magnan, et al. Changing fitness effects of mutations through long-term bacterial evolution. *Science*, 383(6681):eadd1417, 2024.

[55] Guri Giaever, Angela M. Chu, Li Ni, et al. Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, 418(6896):387–391, 2002.

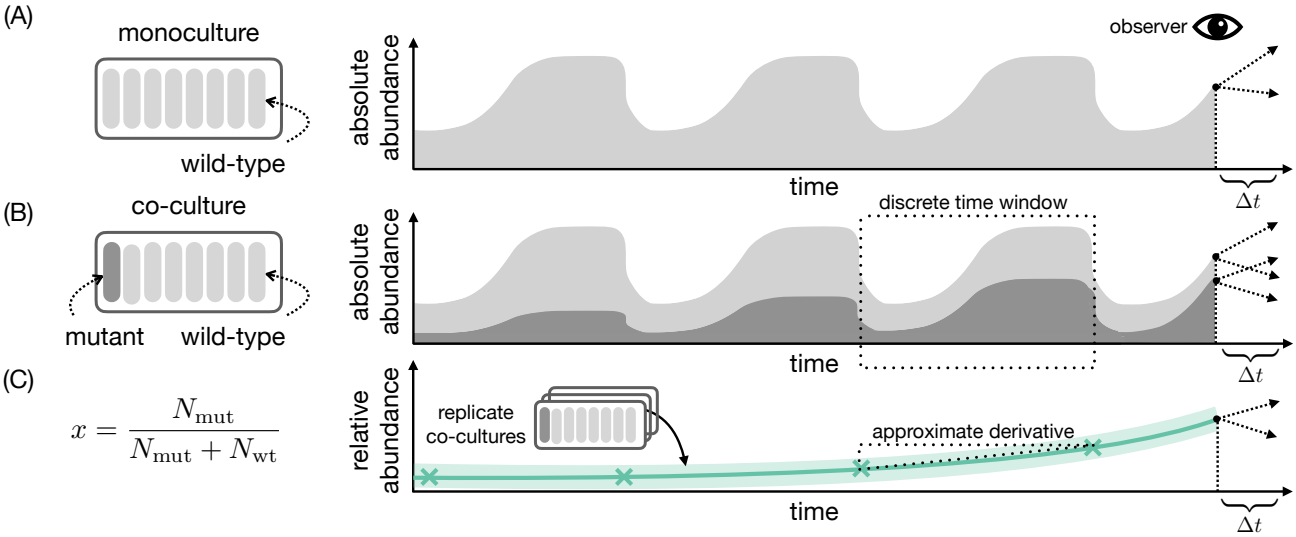

FIG. S1. **Predicting the absolute and relative abundance of microbial populations.** (A) Example timeseries of absolute abundance for a single microbial population (light grey). The observer (eye symbol) of the timeseries at time t can ask about the future trend in absolute abundance (arrows). (B) Example timeseries of absolute abundance for two microbial strains, that may represent a population of wild-type and mutant. The absolute abundance of the wild-type strain (light grey) is stacked on top of the absolute abundance of the mutant strain (dark grey). Similarly, the observer (eye symbol) can ask about the future trend for the absolute abundance of the mutant population or the wild-type population in this co-culture. (C) The relative abundance for the mutant strain corresponding to panel (B). We sketch the mean relative abundance x over time (green line) inferred from multiple replicates (rectangles). An error band (light green) around the timeseries demonstrates the variation between replicates. The observer (eye symbol) can ask about the future trend for the relative abundance of the mutant population (arrows).

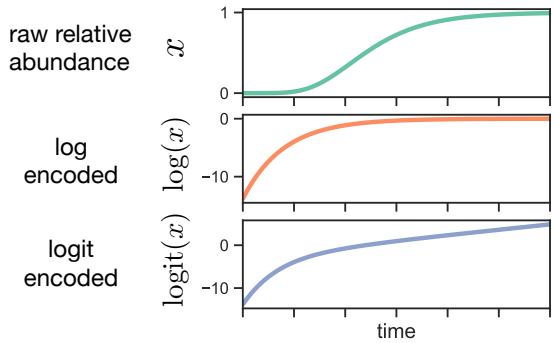

FIG. S2. **The effect of encodings on a non-logistic relative abundance trajectory.** Example rajectory of relative abundance x (top panel) for a mutant invading and eventually replacing a wild-type population, simulated with the Gompertz equation $\frac{dx}{dt} = rx\log(\frac{1}{x})$. Below, we show the Gompertz trajectory under the encoding $\log x$ and the logit-encoding $\log(x/(1-x))$. Compare to Fig. 1A for a trajectory simulated with the logistic equation.

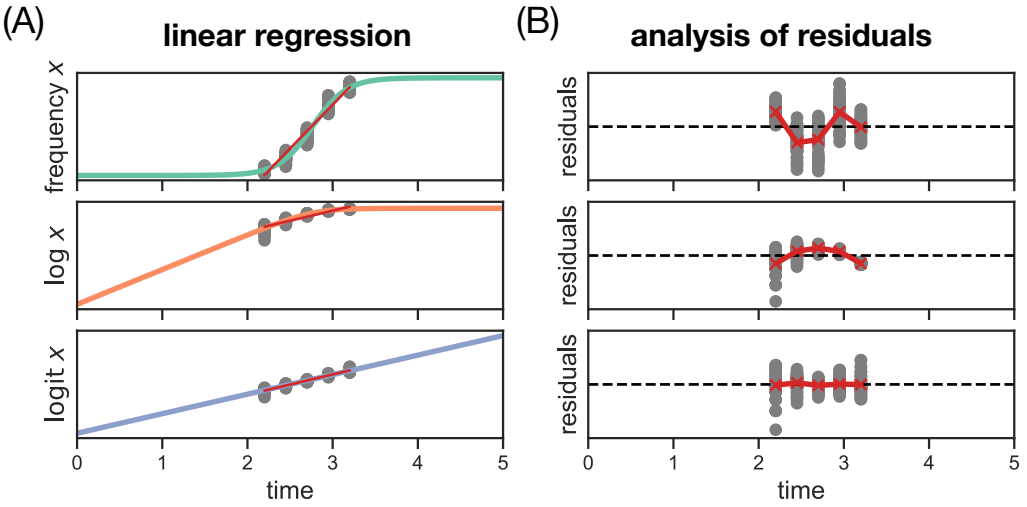

FIG. S3. **The advantages of the logit-encoding for linear regression to relative abundance time-series data.** (A) Linear regression to the example trajectory for mutant relative abundance from Fig. 1A. Given the true relative abundance $x(t)$ (green line), we sample the relative abundance at a set of intermediate timepoints (grey dots; top row) assuming a binomial distribution and fit a linear regression line (red line; top row). We transform samples of raw relative abundance x (grey dots; top row) into the log-encoded abundance $\log x$ (grey dots; middle row) and fit a linear regression (red line; middle row). Similarly, we fit a regression the logit-encoded abundance $\operatorname{logit} x$ (bottom row). (B) Corresponding residuals plot for the regressions in panel A. The sampled relative abundance values (grey dots; panel A) are compared to the fitted regression (red line; panel A) to calculate the residuals (grey dots; this panel). A red line traces the mean value of the residuals. We observe that the residuals under the logit-encoding (bottom row) show no systematic trend in the mean residuals (red line) and have constant variation across timepoints (grey dots).

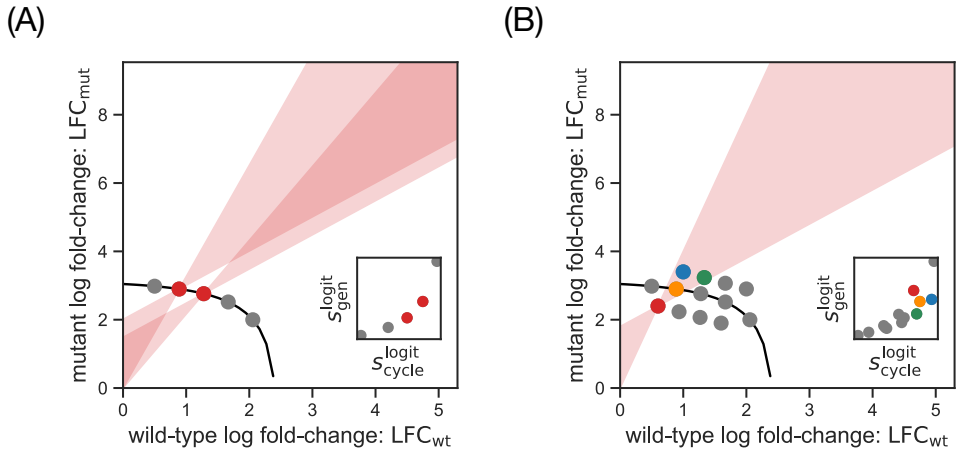

FIG. S4. **The variation of wild-type and mutant log fold-change under resource consumption constraints.** (A) Schematic fold-change variation under a perfect resource consumption constraint between wild-type and mutant. Each dot corresponds to a mutant strain in a 50:50 competition growth cycle with the wild-type strain (Sec. S5). The wild-type LFC is lower in competition with mutants that have very high LFC (red dots) because more resources are consumed by mutant cells. For our model of population dynamics, we can calculate this constraint exactly (black line: Eq. (S36)). We highlight the bow-tie area of misranking (red shading; compare Fig. 1D) for two mutants and show the correlation between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{gen}}$ in the inset. (B) Schematic fold-change variation that deviates from the resource consumption constraint. Additional variation between the competition growth cycles, e.g., due to variation in the biomass yield of the mutant strains, means that the fold-change values do not fall on a single resource consumption constraint (black line; same as in panel A). This can give rise to rank differences in relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{gen}}$ (inset) for those mutants (orange, blue and green dots) that fall in the bow-tie area of the focal strain (red dot).

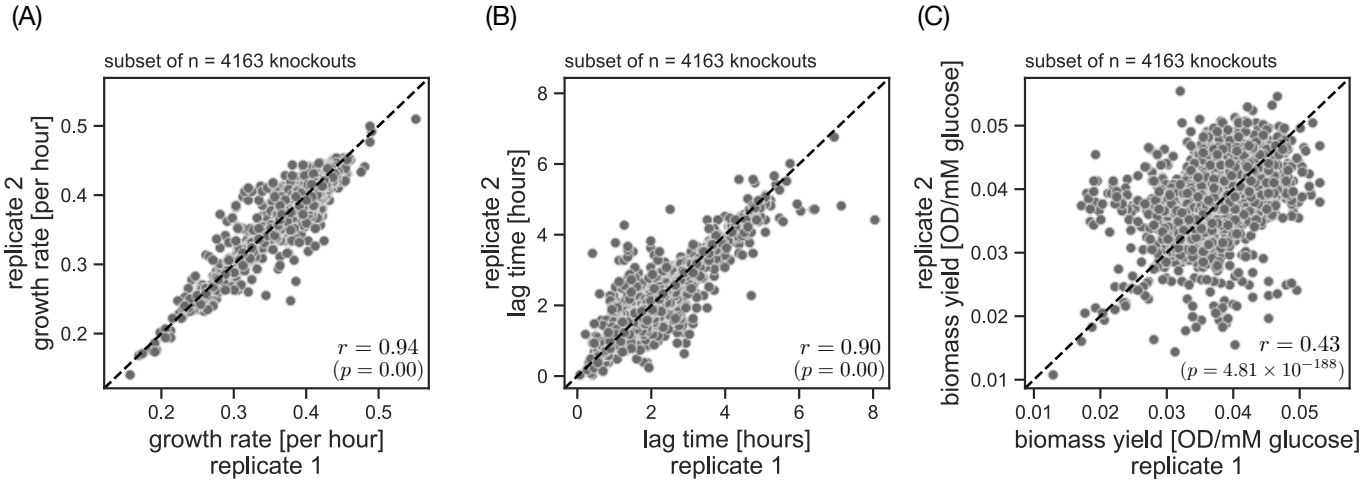

FIG. S5. **Replicate measurements for growth rate, lag time and yield in our empirical trait dataset.** (A) Covariation of growth rate between replicate measurements of the knockouts (grey dots; Pearson correlation coefficient $r = 0.94$, $p = 0.00$). Each dot represents a mutant genotype from the single-gene knockout collection in *Saccharomyces cerevisiae* [55] as measured by Warringer et al. [24]. For the vast majority of genotypes in our dataset ($n = 4163$ out of $n = 4492$ knockouts) we are able to fit two traits from two independent growth curve measurements (Fig. 2A; Methods). (B) Covariation of lag time between replicate measurements of the knockouts (grey dots; $r = 0.90$, $p = 0.00$) (C) Covariation of biomass yield between replicate measurements of the knockouts (grey dots; $r = 0.43$, $p = 4.81 \times 10^{-188}$).

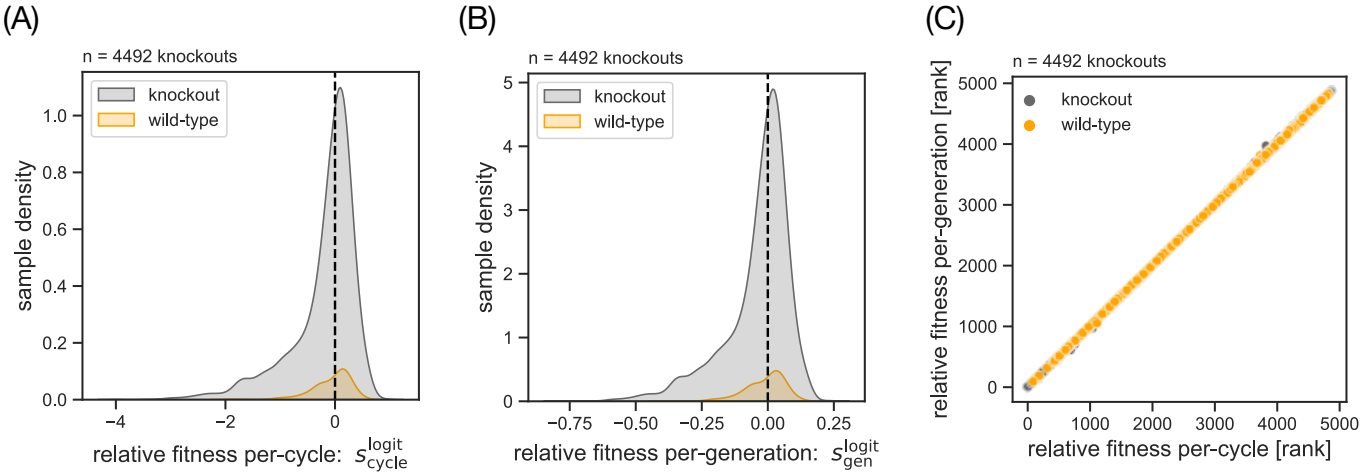

FIG. S6. **The distribution of fitness effects under relative fitness per-cycle and per-generation.** (A) Distribution of relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ for the knockouts (grey) and wild-type replicates (orange) in our dataset (Fig. 2A), based on the fitness data from Fig. 2E. (B) Distribution of relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$. Based on the fitness data in Fig. 2E. (C) Covariation in fitness ranks between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{gen}}$. For each fitness statistic, we calculate the mutant ranking (higher rank means higher fitness and mutants with equal fitness are assigned the lowest rank in the group), based on the rank data from Fig. 2F.

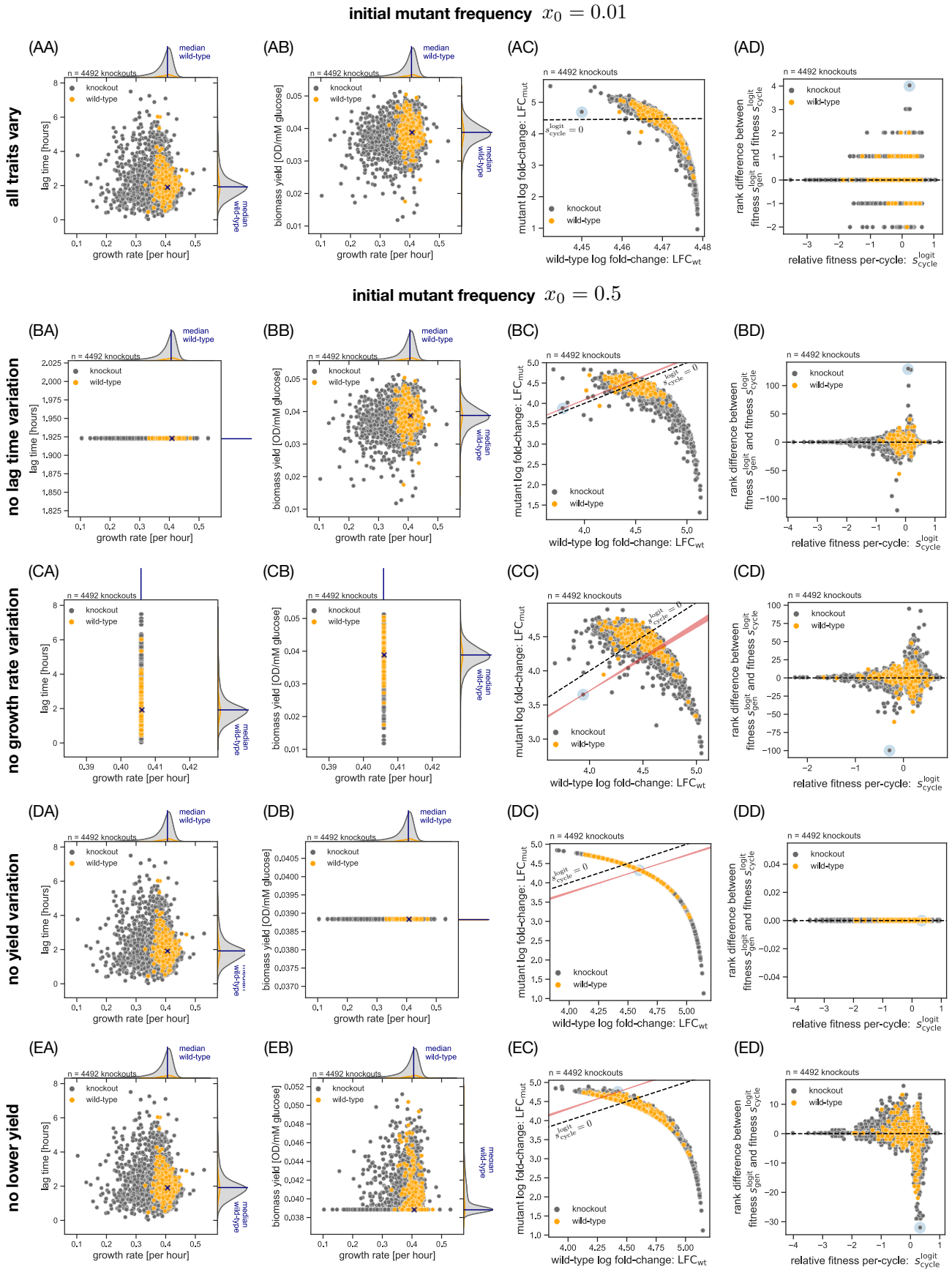

FIG. S7. **Exploring alternative conditions for misranking between fitness statistics in yeast gene-deletion data.** (A: top row) Small mutant frequency $x = 0.01$. (B: second row) Standard mutant frequency $x = 0.5$, but no variation in lag time. (C: third row) Standard mutant frequency $x = 0.5$, but no variation in growth rate. (D: fourth row) Standard mutant frequency $x = 0.5$, but no variation in biomass yield. (E: bottom row) Standard mutant frequency $x = 0.5$, but only positive variation in biomass yield. See Fig. 2 for the default case and panel legends.

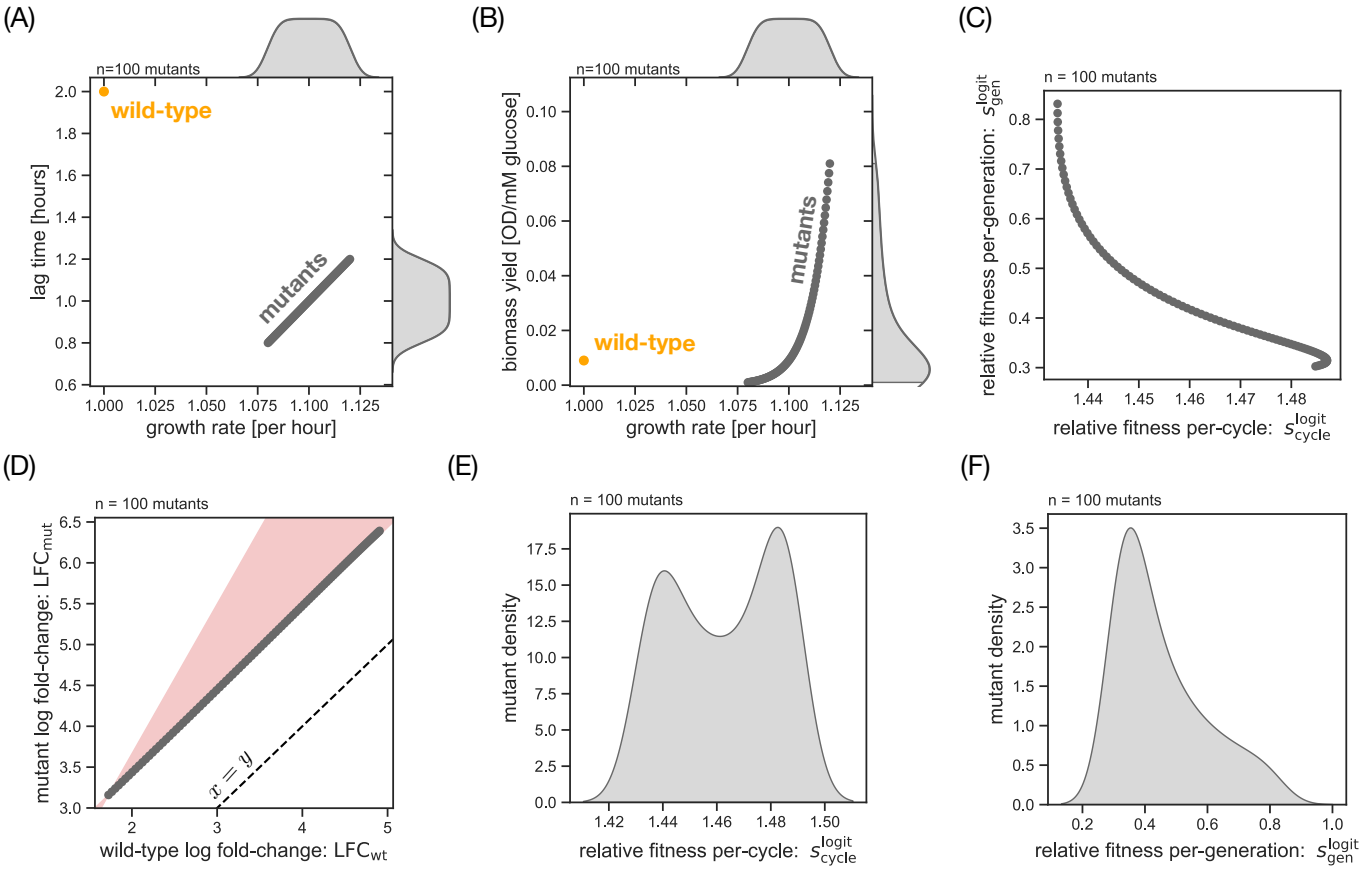

FIG. S8. **Example dataset with anti-correlation between relative fitness per-cycle and per-generation.** (A) Covariation between growth rate and lag time for a synthetically generated dataset of mutants (grey dots) and a single wild-type (orange dots). The variation here may represent the standing variation in an evolved population with improved growth rate and lag time over the wild-type ancestor. The graphic overlap between points means they appear as a single line. (B) Covariation between growth rate and biomass yield for the mutants (grey dots) and wild-type (orange dot) in this synthetic dataset. The variation here is specifically chosen to generate anticorrelation between relative fitnes per-cycle and per-generation. (C) Covariation between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{cycle}}$ for the mutants in panel A-B. We simulate pairwise competitions for all mutants against the wild-type with the exact same settings as for the empirical dataset (Fig. 2D; Methods). (D) Covariation between mutant and wild-type fold-change in the competitions. Based on the competition data in panel C (grey dots), we estimate the LFC values for all mutant-to-wild-type competitions (grey dots). We highlight the bow-tie area of misranking for the mutant with the lowest overall fold-change (red shading; compare Fig. 1D). A dashed black line shows the isocline for zero relative fitness per-cycle (Eq. (8)). (E) Distribution of relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ for the mutants in panel A-B. Based on the fitness data from panel C. (F) Distribution of relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$ for the mutants in panel A-B. Based on the fitness data from panel C.

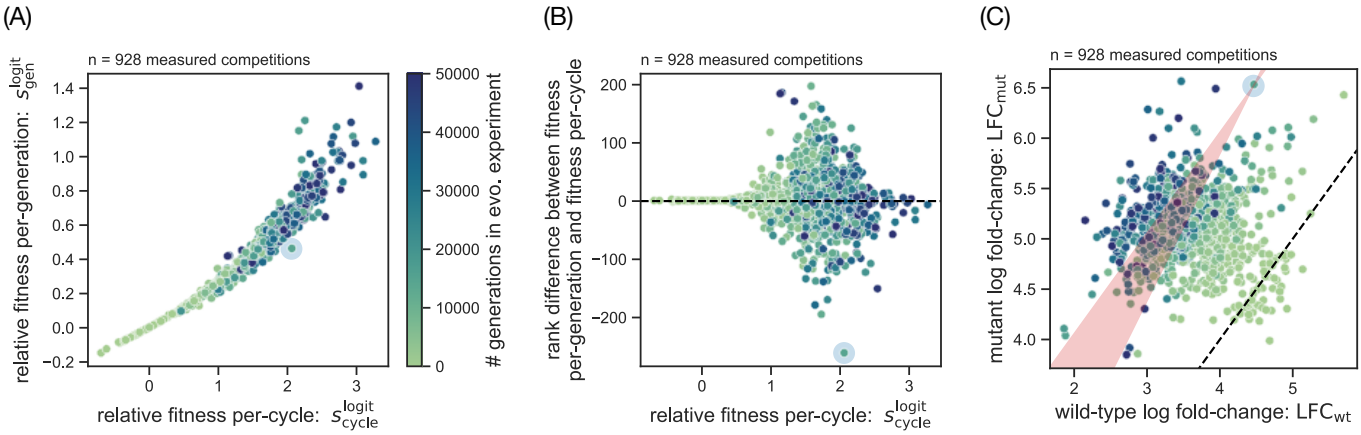

FIG. S9. **Comparison of mutant fitness rankings across the complete LTEE competition dataset.** (A) Covariation between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{gen}}$ for evolved populations of the LTEE. We calculate the fitness statistics based on the competition data published by Wiser et al. [29], who measured the fitness every 250-500 generations (color bar). Each line of the LTEE contributes roughly two competitions per time-point due to replicate measurements [29] (Sec. S8). (B) Rank disagreement between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{gen}}$ for evolved populations of the LTEE. For each fitness statistic, we calculate a ranking (higher rank means higher fitness and mutants with equal fitness are assigned the lowest rank in the group; compare Fig. 2F). The rank difference is defined as the rank in $s^{\text{logit}}_{\text{gen}}$ minus the rank in $s^{\text{logit}}_{\text{cycle}}$. We highlight the evolved population with the greatest rank difference (blue halo). (C) Covariation between the wild-type and mutant fold-change for the competition data from the LTEE [29]. The term 'mutant' (y-axis) refers to an evolved populations at a given time-point from one of the 12 lines of the LTEE. We highlight the evolved population with the greatest rank difference (blue halo; compare panel B) and draw its corresponding bow-tie area of misranking between $s^{\text{logit}}_{\text{cycle}}$ and $s^{\text{logit}}_{\text{gen}}$ (red shading; compare Fig. 1F). The coloring in all panels refers to the time-point at which the evolved population was sampled from the LTEE.

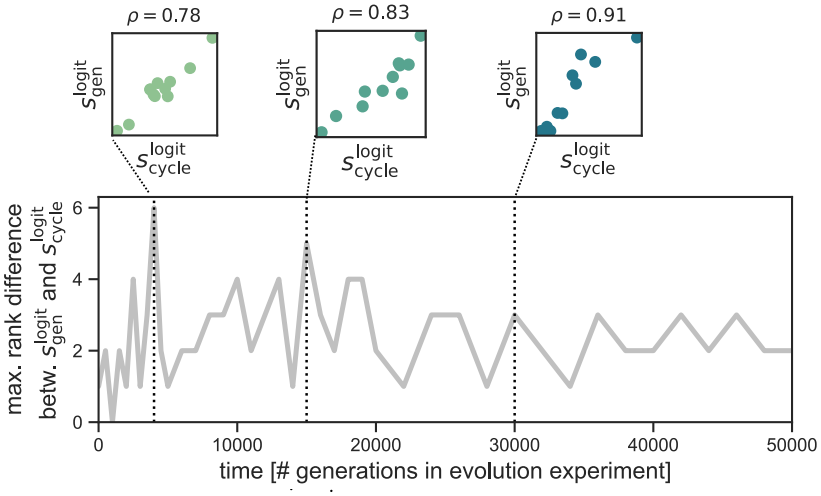

FIG. S10. **Rank difference between relative fitness per-cycle and per-generation as a function of time in the LTEE.** For each of the 12 lines in the LTEE, we compute the relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and per-generation $s^{\text{logit}}_{\text{gen}}$ as a function of time, by averaging the data from the LTEE competition dataset [29] across replicate measurements (Sec. S8). For some lines, the time-series is truncated due to measurement difficulties [29]. Based on the quantitative fitness values $s^{\text{logit}}_{\text{cycle}}$ and $s^{\text{logit}}_{\text{gen}}$, we compute fitness rankings between the replicate lines at any given timepoint (higher rank means higher fitness and mutants with equal fitness are assigned the lowest rank in the group; compare Fig. 2F). The rank difference is defined as the rank in $s^{\text{logit}}_{\text{gen}}$ minus the rank in $s^{\text{logit}}_{\text{cycle}}$, and we combine the rank difference between all pairs of evolving lines to compute the maximum rank difference at each time-point. At three chosen time-points ($t = 4000$, $t = 15000$, $t = 30000$ generations), we show the correlation between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$ (each dot corresponds to one of the 12 lines in the LTEE). We quantify the correlation between $s^{\text{logit}}_{\text{cycle}}$ and $s^{\text{logit}}_{\text{gen}}$ with the Spearman rank correlation ρ (panel title). Colors indicate the time-point and correspond to the color bar in Fig. S9.

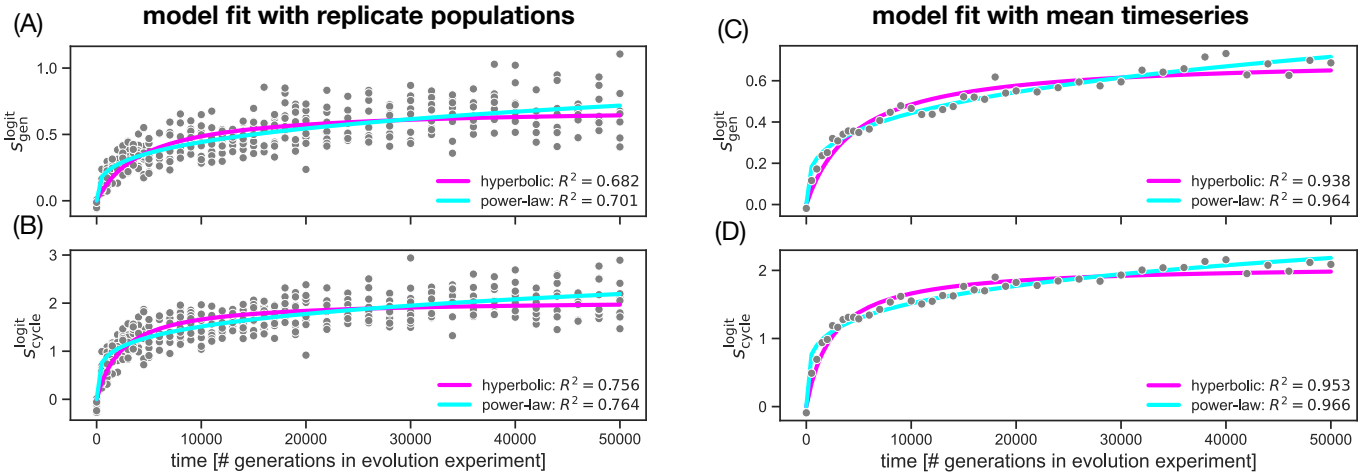

FIG. S11. **Long-term fitness trends in the LTEE under relative fitness per-cycle and per-generation.** (A) Fit of a hyperbolic model (pink line; Eq. (S45)) and a power-law model (cyan line; Eq. (S46)) to a pooled time-series of relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$. We pool the fitness per-generation $s^{\text{logit}}_{\text{gen}}$ from all 12 lines of the LTEE into a single dataset (grey dots) and repeat the fits of Wiser et al [29] (Sec. S8). We compute the fraction of variance explained R^2 as a measure for the quality of fit (hyperbolic model: $R^2 = 0.682$, power-law model: $R^2 = 0.701$). (B) Fit of hyperbolic (pink) and power-law models (cyan) to a pooled time-series of relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$. Similar to panel A, we pool the relative fitness-per cycle $s^{\text{logit}}_{\text{cycle}}$ from all 12 lines of the LTEE into a single dataset (grey dots) and identify a the best fit for the hyperbolic model ($R^2 = 0.756$) and and the power-law model ($R^2 = 0.764$). (C) Fit of hyperbolic (pink) and power-law models (cyan) to an averaged time-series of relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$. Following the 'grand-mean' averaging strategy outlined in Wiser et al. [29], we compute the average relative fitness per-generation $s^{\text{logit}}_{\text{gen}}$ across all evolving lines (grey dots; Methods) and identify a the best fit for the hyperbolic model ($R^2 = 0.938$) and the power-law model ($R^2 = 0.964$). Note that the fraction of variance explained in is much higher compared to panel A, because the averaged time-series has fewer points (Sec. S8). (D) Fit of hyperbolic (pink) and power-law models (cyan) to an averaged time-series of relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$. Simlar to panel C, we compute the average relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ across all evolving lines in the LTEE and identify the best fit for the hyperbolic model ($R^2 = 0.953$) and the power-law model ($R^2 = 0.966$). For a comparison between our results and the original fit by Wiser et al [29], see Sec. S8.

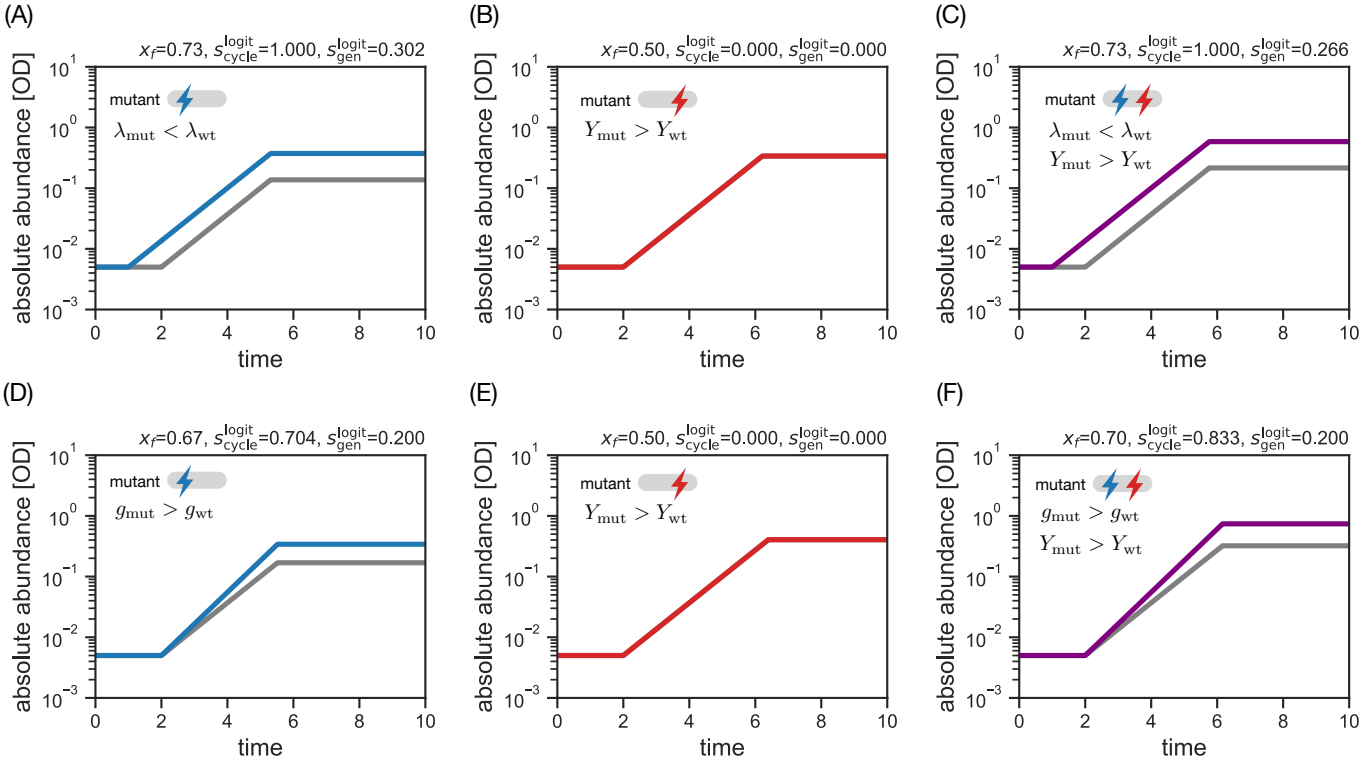

FIG. S12. **Comparing magnitude epistasis between relative fitness per-cycle and per-generation.** (A) Simulated competition growth cycle for a single mutant with short lag time (blue line) in co-culture with a wild-type strain (grey line). We simulate the co-culture using the same settings as for our empirical dataset (Fig. 2D; Methods). In the top, we show the final relative abundance of the mutant x_f, the relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ (Eq.(8)) and per-generation $s^{\text{logit}}_{\text{gen}}$ (Eq. (9)). (B) Simulated competition growth cycle for a single mutant with higher biomass yield (red line), competing against the same wild-type as in panel A (grey line) (C) Simulated competition growth cycle for a double mutant (purple line) with shorter lag time (as in panel A) and higher biomass yield (as in panel B). Compare panel A-C to Fig. 3C-D for the epsistasis plot. (D) Simulated competition growth cycle for single mutant with higher growth rate (blue line). (E) Simulated competition growth cycle for a single mutant with higher biomass yield (red line), same as panel B. (F) Simulated competition growth cycle for a double mutant (purple line) with higher growth rate (as in panel D) and higher biomass yield (as in panel E). Compare panel D-F to Fig. 3E-F for the epistasis plot.

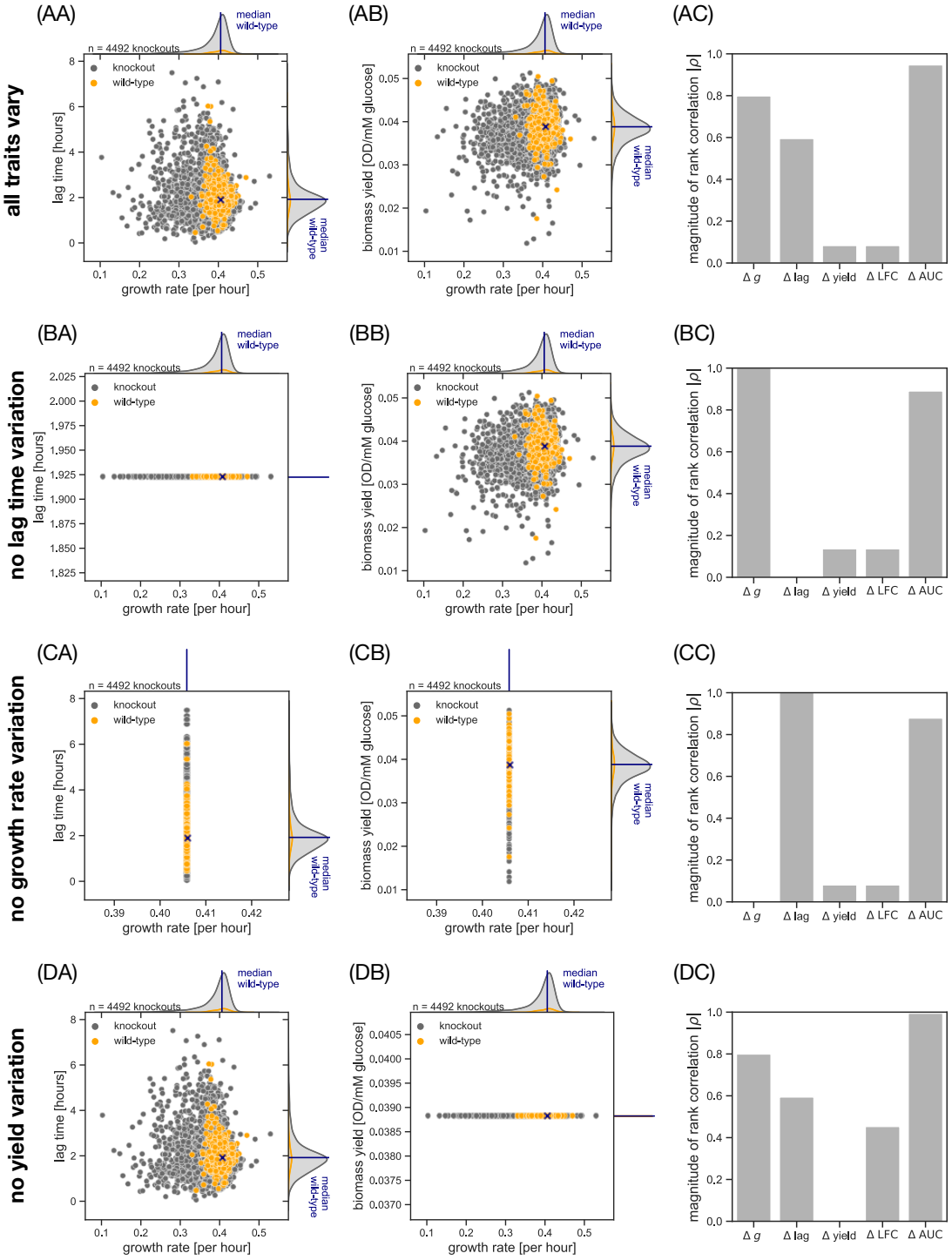

FIG. S13. **Predicting relative fitness with monoculture proxies under different scenarios of trait variation.** (AA) Covariation between growth rate and lag time in our empirical trait distribution (same as Fig. 2B). (AB) Covariation between growth rate and biomass yield in our empirical trait distribution (same as Fig. 2C). (AC) Quality of prediction for different monoculture proxies under the trait distribution in panel AA-AB. As the ground truth, we estimate the relative fitness per-cycle $s_{\text{cycle}}^{\text{logit}}$ using a simulated 50:50 competition growth cycle (Fig. 2D; Methods). For each genotype i, we compute the growth rate difference $\Delta g = g_i - g_{\text{wt}}$, the lag time difference $\Delta\text{lag} = \lambda_i - \lambda_{\text{wt}}$ and the difference in biomass yield $\Delta\text{yield} = Y_i - Y_{\text{wt}}$. Additionally, we compute the difference in log fold-change $\Delta\text{LFC} = \text{LFC}_i - \text{LFC}_{\text{wt}}$ and the difference in area under the curve $\Delta\text{AUC} = \text{AUC}_i - \text{AUC}_{\text{wt}}$ from simulated monoculture growth curves (Sec. S9). We quantify the agreement between the monoculture proxy and the relative fitness per-cycle $s_{\text{cycle}}^{\text{logit}}$ using theSpearman correlation coefficient ρ, which reflects the agreement in ranking. (B: second row) Modified trait distribution with no variation in lag time. (C: third row) Modified trait distribution with no variation in growth rate. (D: fourth row) Modified trait distribution with no variation in biomass yield.

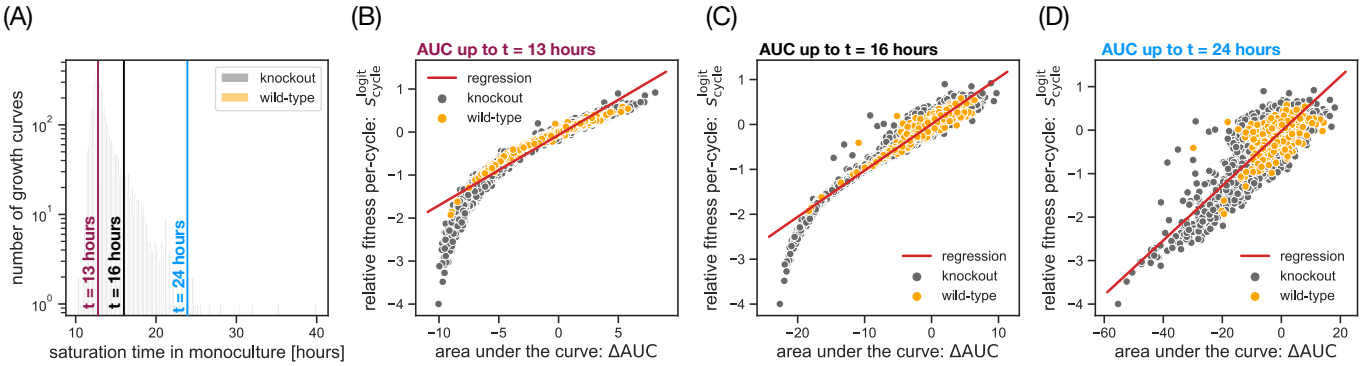

FIG. S14. **The choice of the cut-off time for evaluating the area under the curve (AUC).** (A) Distribution of saturation times in mono-culture for the knockouts (grey bars) and wild-type replicates (orange bars; not visible) in our empirical dataset (Fig. 2A). The saturation time t_{sat} is defined as the time when the limiting resource is depleted ($R(t_{\text{sat}}) = 0$) and can be estimated numerically from the trait data (Methods). Three vertical lines indicate example choices for the cut-off time t_{eval} of the AUC: the most frequent saturation time ($t_{\text{eval}} = 13$ hours; purple line), an external timescale ($t_{\text{eval}} = 24$ hours; blue line) and as saturation time half-way in the decay of the distribution ($t_{\text{eval}} = 16$ hours; black line). (B) Covariation between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and AUC with a short cut-off time. We compute the AUC from a simulated monoculture knockouts (grey dots) and wild-type replicates (orange dots) using the cutoff time $t_{\text{eval}} = 13$ hours (Sec. S9). As the ground truth, we take the relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ from a pairwise competition (Fig. 2A). As a red line, we show the best fit from a linear regression of $s^{\text{logit}}_{\text{cycle}}$ to the AUC. (C) Covariation between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and AUC with an intermediate cut-off time ($t_{\text{eval}} = 16$ hours). (D) Covariation between relative fitness per-cycle $s^{\text{logit}}_{\text{cycle}}$ and AUC with a longer cut-off time ($t_{\text{eval}} = 24$ hours).

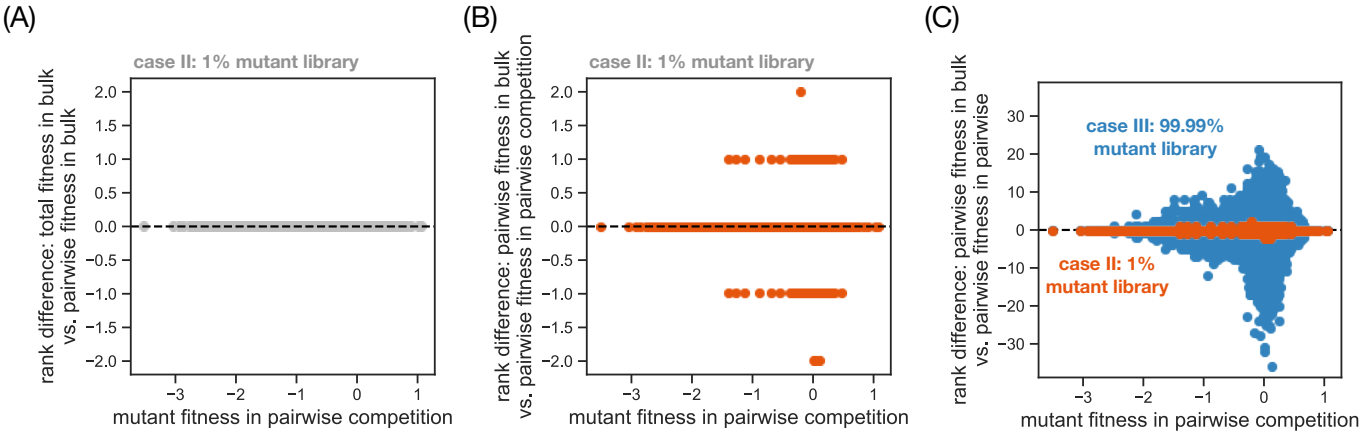

FIG. S15. **The error in mutant fitness rankings between bulk and pairwise competition experiments.** (A) Rank difference between total relative fitness and pairwise relative fitness in a bulk competition growth cycle with low mutant library abundance (Fig. 4A, case II). Based on the fitness data in Fig. 4C, we calculate a mutant ranking for total relative fitness in bulk (Eq. (22)) and a ranking for pairwise relative fitness in bulk (Eq. (23)) (higher rank means higher fitness and mutants with equal fitness are assigned the lowest rank in the group). The rank difference is defined as the rank in total relative fitness $\sigma^{\text{logit}}_{i,\text{cycle}}$ and minus the rank in pairwise relative fitness $s^{\text{logit}}_{i,\text{wt},\text{cycle}}$. (B) Rank difference between pairwise relative fitness in bulk (Eq. (23)) and the relative fitness per-cycle in a pairwise competition (Eq. (8)). Based on the pairwise relative fitness at low mutant library abundance in Fig. 4C (red dots; case II), we calcuate a mutant ranking for pairwise relative fitness in bulk (Eq. (23)) and a mutant ranking for the relative fitness per-cycle in pairwise competition (Eq. (8). The rank difference (red dots) is defined as the rank in bulk competition minus the rank in pairwise competition. (C) Rank difference between pairwise relative fitness in bulk and the relative fitness per-cycle in a pairwise competition for two different cases of the bulk competition. Based on the fitness data in the inset of Fig. 4C, we estimate a mutant ranking for pairwise relative fitness (Eq. (23)) in the case of a bulk competition growth cycle with high mutant library abundance (blue dots; compare Fig. 4A, case III). For each case, the rank difference is defined as the rank in the bulk competition minus the rank in pairwise competition and the rank difference for case II (red dots) is identical to the data in panel B. All fitness statistics in this figure are based on the logit-encoding, however, since the underlying relative abundances are small ($x \ll 1$), it is approximately equivalent to the log-encoding ($\log x \approx \text{logit}\, x$).

FIG. S16. **A decomposition for the error from higher-order interactions in bulk competition experiments.** For a bulk competition growth cycle with low mutant library abundance (Fig. 4A, case II), we calculate the pairwise relative fitness (Eq. (23)) for the knockouts in our empirical dataset (Fig. 2A) using a previously established approximation [28] (Sec. S13). The error from higher-order interactions is defined as the pairwise fitness in bulk minus the fitness in pairwise competition. Based on the approximation for our model of population dynamics (Sec. (S13)), we derive a decomposition that separates the absolute error into two terms that we call the *fitness-dependent error term* (dark dots) and the *fitness-independent error term* (light dots). For full details on the decomposition, see Sec. (S14).

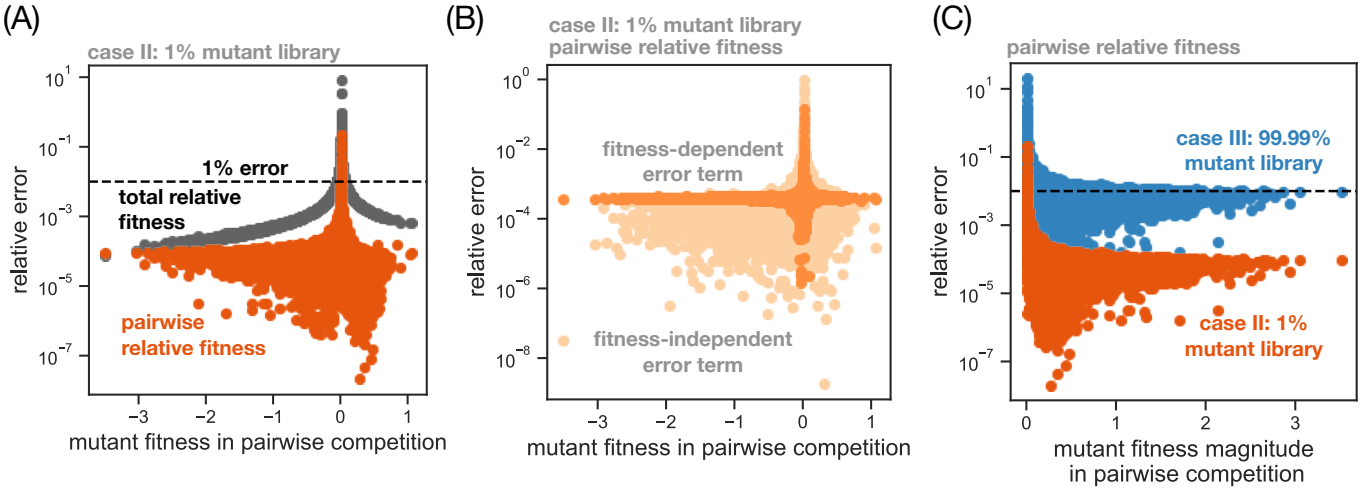

FIG. S17. **The relative error between bulk and pairwise competition experiments.** (A) Comparing the relative error between total and pairwise relative fitness in bulk. Based on the absolute error in Fig. 4C, we estimate a relative error for the total relative fitness (grey dots; Eq. (22)) and pairwise relative fitness (red dots; Eq. (23)) in a bulk competition growth cycle with low mutant library abundance (Fig. 4A, case II). The relative error of each bulk statistic is defined as the absolute error (Fig. 4C), divided by the fitness in pairwise competition (Eq. (S91)). A dashed grey line indicates the threshold of 1% relative error. (B) Comparing the relative error between the components of higher-order interactions. Based on the absolute error from higher-order interactions in Fig. S16, we estimate the relative error for the fitness-dependent (dark dots) and the fitness-independent error component (light dots). Here the relative error is defined as the absolute error component (Fig. S16), divided by the relative fitness per-cycle in pairwise competition (see Sec. S14 for details). (C) Comparing the relative error between low and high mutant library abundance in bulk. Based on the absolute error in the inset of Fig. 4C, we estimate a relative error for the pairwise relative fitness at low mutant library abundance (red dots; Fig. 4A, case II) and high mutant library abundance (blue dots; Fig. 4A, case III). The relative error for each case is defined as the as absolute error (Fig. 4C; inset), divived by the fitness in pairwise competition (Eq. (S91)). On the x-axis, we plot the absolute value of relative fitness per-cycle in the pairwise competition.