Fitness landscape analysis reveals that the wild type allele is sub-optimal and mutationally robust
- Supporting Information

Tzahi Gabzi, Yitzhak Pilpel and Tamar Friedlander

September 25, 2021

1 Assessment of fitness measurement errors

The fitness values were calculated in [1] by sequencing culture samples at two time points, $T = 0$, $T = 24$ h and estimating the relative frequencies of the different genotypes in the culture. Fitness calculation assumed that the cells grew exponentially between these two time points (See detailed description in the Methods section, main text). Our focus in this paper was on the fraction of genotypes with fitness higher than the wild type's. Below, we scrutinize potential sources of error in the fitness value estimate, in order to rule out the possibility that these high-fitness values are due to an experimental artifact.

1.1 Exponential growth assumption

The underlying assumption in the fitness calculation is that all genotypes grew exponentially during the entire experiment (24 h.). In practice, before entering exponential growth, cells spend time in a preparatory phase called a 'lag phase' (in which growth rate is relatively slow). The length of the lag phase could vary between strains, hence sampling all genotypes at the same time could potentially catch some of them still in the lag phase. In a batch culture, once the cells use up the available nutrients, their numbers saturate and growth slows down again (also called 'yield' phase). Although the culture was diluted after 12 h [1] some strains could have reached their yield phase earlier than others. As we do not have detailed growth curves, it is hard to assess how common this is and to what extent it could have affected the calculated fitness values. However, this error source, if it exists, is expected to only *underestimate* fitness values. That is because, the final genotype frequency could have been attained in a shorter time period than assumed in the calculation, and thus the growth rate must have been larger than estimated.

1.2 Read-count noise

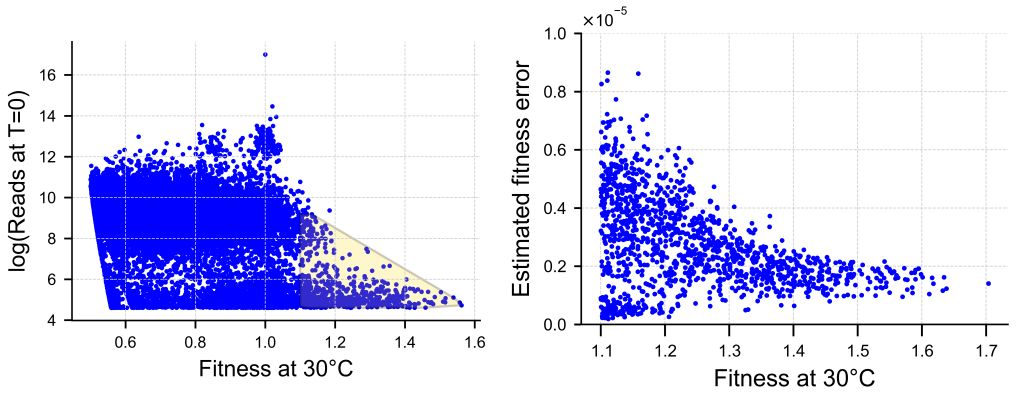
The frequencies of the different genotypes in the culture were estimated using samples of the batch culture at two time points where only genotypes with at least 100 reads at $T = 0$ were considered [2]. Sampling errors could potentially lead to error in fitness estimates. In contrast to the previous error source, which could only cause under-estimation of fitness values, read-count noise could cause both over and under-estimation of fitness. A combination of under-sampling at $T = 0$ and over-sampling at $T = 24$ h of a particular genotype could cause an *over-estimate* of its fitness. In contrast, over-sampling at $T = 0$ and under-sampling at $T = 24$ h should lead to the opposite effect of fitness value *under-estimation*. Sampling errors in $T = 0$ and $T = 24$ h could also (at least partially) cancel each other, if they are both in

the same direction. In Fig. S1a, we show the number of reads at $T = 0$ against the calculated fitness values. We are mostly concerned with presumably high-fitness genotypes that had a low number of reads at $T = 0$, and hence their fitness could have been over-estimated. To estimate the errors, we used the number of reads obtained in the experiments for each genotype. A common sampling of all genotypes at $T = 0$ was applied. This common culture was used to inoculate the multiple growth experiments under the four different growth conditions, with 3-5 independent replicates under each condition. Thus, at $T = 24$ h, there were multiple samples for each genotype under each condition (one for each replicate). We estimated these sampling errors by the standard deviations of the associated binomial distributions, following the procedure introduced by Levy *et al.* [3]. For the error analysis, we took the two extremes: To estimate the extent to which the fitness could be *underestimated*, we added the standard deviation to the value measured at $T = 0$ and subtracted the corresponding standard deviation at $T = 24$ h. To calculate the extent to which the fitness could be *overestimated* we did the opposite, by subtracting the standard deviation at $T = 0$ and adding it at $T = 24$ h. As the reported fitness values were averaged over the multiple replicates, here too, we averaged the fitness overestimates (at all replicates) and similarly averaged all fitness underestimates. Differences of the over- and underestimates from the reported fitness values are the two fitness deviations for each genotype. We then take the maximum over these two deviations (absolute-valued) to represent the typical error in the fitness value. In Fig. S1b, we plot this error (calculated as explained above) relative to the reported fitness value, for each genotype with fitness higher than 1.1. We find that the relative fitness errors are all smaller than 1×10^{-5}. Hence, the many fitness values higher than 1.1 cannot all be dismissed as experimental error.

The fitness values reported by Li *et al.* are averages of the fitness values calculated from the distinct biological replicates. As an additional estimate of the fitness values inaccuracy, we took the fitness values calculated from single-replicate experiments and calculated the standard deviation over these replicates - see Fig. S2. Differences in fitness estimates between replicates could be due to read count as well as due to other biological factors. Unsurprisingly, the standard deviation between the biological repeats yielded higher fitness errors than our previous estimate of the inaccuracies due to read-count alone. In either estimate though, only a small fraction of the genotypes found in the experiment to be fitter than the wild type, can be attributed to measurement inaccuracy.

1.3 Background mutations

Mutations could have also occurred in other genome locations during the course of the competition experiment. While not a measurement error, had such a mutation occurred, we could have mistakenly attributed the fitness effect to the tRNA mutation. The cells barcode identifies them by their tRNA variant. If such a background mutation occurred during the course of the experiments, after cells had already started reproducing, it would result in cells carrying a common barcode that are a mixture of multiple sub-populations: the original genome and the one carrying the background mutations. The later such a mutation emerges, the smaller its population fraction. Since sequencing cannot distinguish between the two types, the estimated fitness would average over the two sub-populations. Thus, the effect of the background mutation would be largest, if it occurred at the very beginning when only a single cell of each tRNA variant existed. We are particularly interested in determining whether high-fitness genetic variants can be explained by such background mutations; below we will



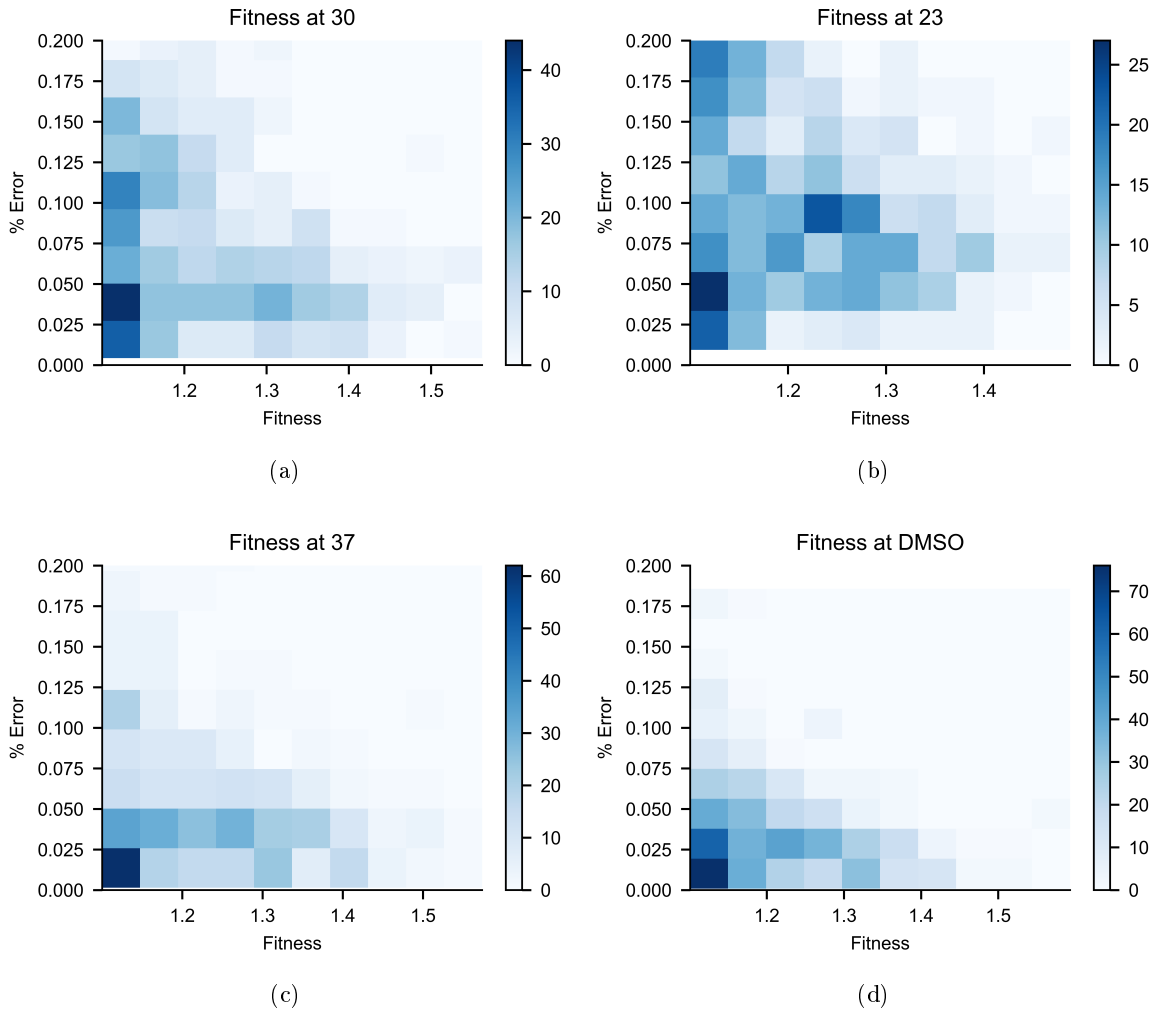
(a) (b)

Figure S1: **Estimation of potential errors in fitness values due to read-count inaccuracy. (A)** Scatter plot of the logarithm of the reported number of reads for each variant at the beginning of the competition ($T = 0$) against the reported fitness values (at 30°C). The region marked in yellow highlights genotypes with reported high-fitness values and low number of initial reads. Low estimate of the number of reads at $T = 0$ could lead to overestimation of the fitness value. **(B)** Estimate of the relative fitness error due to read count noise at both $T = 0$ and $T = 24$ h for each genotype with reported fitness > 1.1 at 30°C.

focus on beneficial mutations alone. Both neutral and deleterious mutations are expected to have only a minor effect on fitness estimates. Neutral mutations have no effect on fitness to begin with and deleterious mutations are naturally selected against and hence will be present in low numbers.

In the experimental procedure used by Li *et al.* , the different variants were synthesized using error-prone PCR initiated with the wild type strain. The products were then transformed into yeast cells (See [1, 4] SI for more details). We estimate that only a small fraction ($\approx 0.001\%$) of the yeast cells was transformed with a variant, and hence each variant exists in very few cells. For simplicity, we assume here that each variant exists as a single copy only, which is the worst-case scenario for our purpose.

To estimate the likelihood of naturally occurring beneficial mutations, we use results from Levy *et al.* [3], where the evolutionary dynamics of 500,000 different yeast lineages, each tagged by a unique barcode, was tracked. They reported the rate of beneficial mutations as a function of their fitness effect. For example, the rate of beneficial mutations that led to fitness increment $s > 5\%$ was approximately 10^{-6} per cell per generation. This is equivalent, in our formalism, to fitness $f = 2^s > 1.03$. In our dataset, there were ∼2000 genotypes with fitness higher than the wild type's and the transformation phase lasted 50 generations. Hence, the probability that such a mutation occurred in one of these ~ 2000 strains in the first generation is only 0.002. If we account for mutations occurring during the entire experiment (50 generations), the probability is larger (0.1), but then these cells are mixed with non-mutated cells carrying the same barcode, such that the fitness effect of the mutations needs to be much larger for it to have the same measured fitness effect when averaged with non-mutated cells. Since the spectrum of beneficial mutation rates reported by Levy *et al.* drops sharply



(a) (b) (c) (d)

Figure S2: **The relative fitness error over multiple biological replicates against the fitness of high-fitness variants at all four conditions.** The relative fitness error - the standard deviation of fitness values obtained at single-replicate experiments divided by the mean - against the reported fitness - shown as 2D histograms. Each panel shows one of the four conditions. While there are $\sim$ 2000 genotypes with fitness > 1, we find that only for a few dozens of them (depending on the conditions), the relative fitness error over multiple biological replicates is larger than 10%.

near $s = 12\%$ fitness benefit, we estimate such a beneficial mutation to be far less likely. This leaves the expectation of background beneficial mutations during transformation at 10^{-3}, far less from the 8%-10% proportion of beneficial mutations found in the Li *et al.* data. Hence, this error source is obviously negligible.

2 Steepness calculation

The fitness landscape was sampled non-uniformly, with dense sampling close to the wild type and sparser sampling further away. Hence, we not only have to calculate genotype steepness using only a subset of their 1-neighbors, but the choice of 1-neighbors is biased to genotypes closer to the wild type. This can potentially affect the steepness estimates, if the fitness landscape is not uncorrelated. To test the sensitivity of the steepness calculation to the non-uniform sampling of the landscape, we compared the steepness values obtained using different subsets of 1-neighbors: far from the wild type (Hamming distance $\geqslant$2 from the wild type), and close to the wild type 1-neighbors (Hamming distance $\leqslant$2 from the wild type). 1-neighbors at distance 2 were a large fraction of the genotypes and hence were included in both sets in order to avoid biases of the calculation. Note that in general, there is a larger total number of 2-, 3-etc. neighbors in the dataset, and we did not normalize for that. We specifically calculated this for fitter-than the wild type mutants, to corroborate our finding that the wild type is amongst the flattest genotypes in the dataset.

We then calculated the steepness for each of those high-fitness genotypes twice: using each of the neighbor subsets - see results in Fig. S3. We found that steepness values calculated using only 1-neighbors close to the wild type (Hamming distance $\leqslant 2$), were somewhat lower (flatter) than steepness values calculated using further 1-neighbors (Hamming distance $\geqslant 2$). The wild type is relatively flat. Hence, its single mutants have fitness values similar to the wild type's fitness, and thus their fitness values are also similar to each other's. The steepness calculation using the close-to-wild-type-1-neighbors includes genotypes from the wild type's 1-neighbors and hence it is reasonable that it yields low steepness values compared to the other calculation. We conclude that there is some bias in the steepness calculation, at least close to the wild type. This also means that there is some level of fitness correlation in this fitness landscape, which is very typical of biological fitness landscapes [5].

Despite this bias, with either calculation, the wild type steepness value remains at the low end of the steepness distribution. As our steepness calculations include more close-to-wild-type than far-from-wild-type 1-neighbors, we are likely underestimating the steepness values of many genotypes. Hence, the steepness difference between the wild type and the other genotypes can be even greater than what we reported, based on the partial set of mutants.

2.1 In the NK model, mean fitness of neutral evolution exactly overlaps with the landscape mean fitness

In Fig. 5a of the main text, we also compared the fitness to the median fitness value of the entire dataset (horizontal blue dashed line). The fitness obtained in the control simulation was very close to the median fitness value, but was consistently slightly higher. If the neutral simulation uniformly sampled all the genotypes in the dataset, we would expect it to overlap with the dataset median fitness value. For comparison, we ran a similar evolutionary simu-

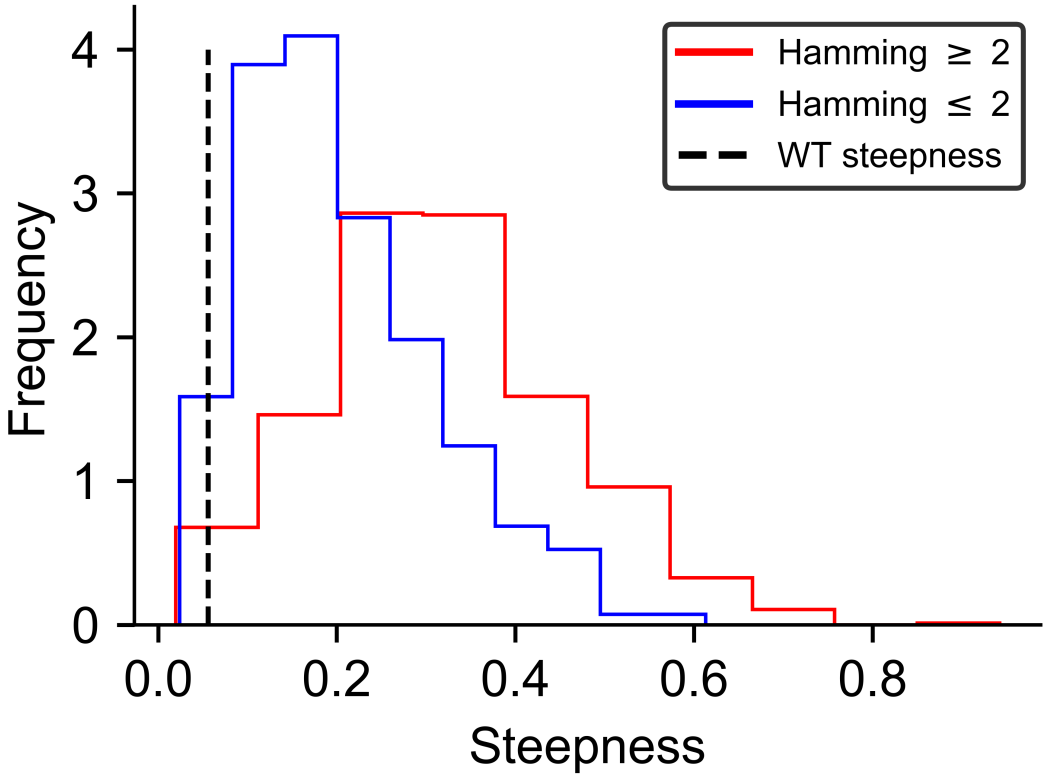
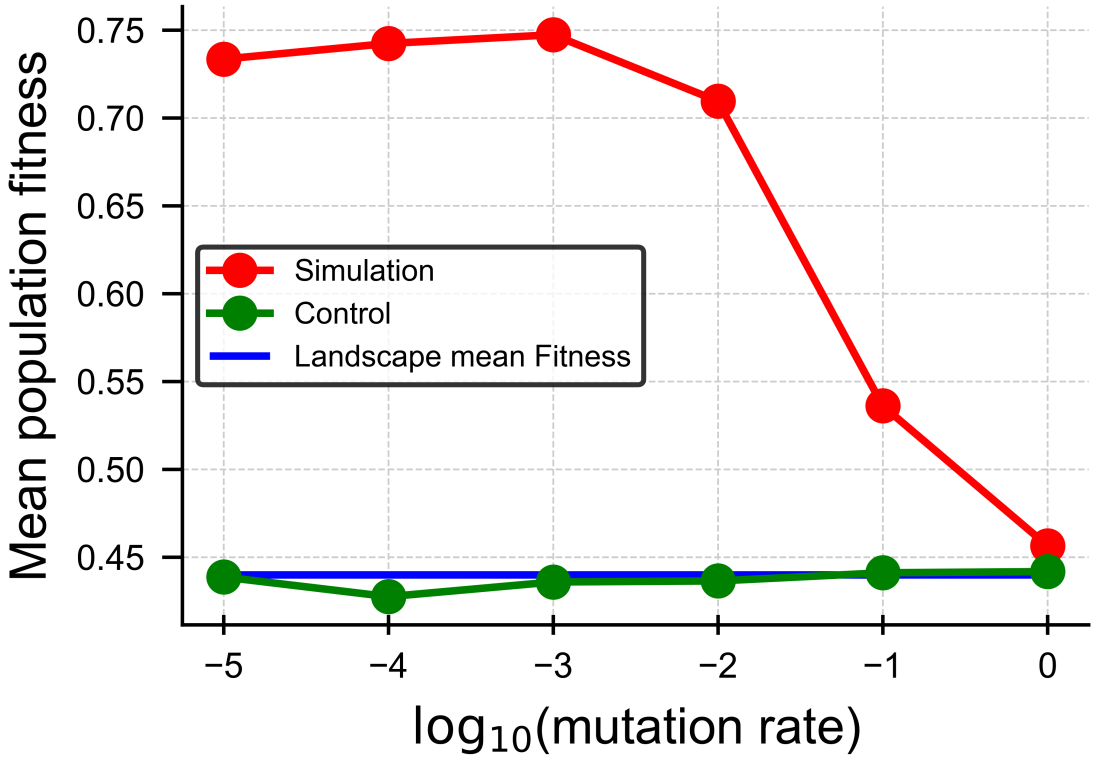


Figure S3: **Comparing steepness calculations using two different subsets of 1-neighbors.** We calculated the steepness values for the same set of fitter-than wild type genotypes using two different sets of 1-neighbors: their 1-neighbors far from the wild type (red curve) and their 1-neighbors close to the wild type (blue curve). We plot here the distributions of steepness values obtained in these calculations. The wild type steepness is shown for reference (vertical black dashed line).

lation with a neutral control on an artificially fabricated NK model landscape for which we have full data of all genotypes and equal connectivity for all genotypes - see Fig. S4. Here, the fitness of the control simulation exactly overlapped with the dataset median fitness value. Hence, we attribute the small gap between the control (green points) and the dataset median found in main text Fig. 5a to some level of correlation between connectivity and fitness in our dataset.



(a)

Figure S4: **NK model: mean population fitness at different mutation rates - simulation results.** We plot here the mean population fitness as obtained in evolutionary simulations, against the mutation rate (red curve). As in the tRNA simulations (main text. Fig. 5), we also ran a control simulation with equal fitness for all genotypes (green curve). We plot, for reference, the mean fitness value of the entire landscape (blue horizontal line). Similarly to the tRNA landscape, we observe a drop in fitness beyond a critical mutation rate (here found at approximately 10^{-2}). As in the evolutionary simulations over the tRNA landscape, here too, the control simulation exhibits no dependence on the mutation rate. Here, however, the mean fitness in the control simulation coincides with the whole landscape mean fitness value. Simulation parameters: $N = 7$, $K = 1$, Population size $= 5000$, random initialization. The simulation ran for 200 generations each time. Each dot is an average of multiple repeats for each parameter setting.

References

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

[2] Shu Li, Justin S. Williams, Penglin Sun, and Teh-hui Kao. All 17 S-locus F-box proteins of the S2- and S3-haplotypes of Petunia inflata are assembled into similar SCF complexes with a specific function in self-incompatibility. *The Plant Journal*, 87(6):606–616, September 2016.

[3] Sasha F. Levy, Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181–186, March 2015. Number: 7542 Publisher: Nature Publishing Group.

[4] Chuan Li and Jianzhi Zhang. Multi-environment fitness landscapes of a tRNA gene. *Nature Ecology & Evolution*, 2(6):1025, June 2018.

[5] Sergey Kryazhimskiy, Gašper Tkačik, and Joshua B. Plotkin. The dynamics of adaptation on correlated fitness landscapes. *Proceedings of the National Academy of Sciences*, 106(44):18638–18643, November 2009.