# Fitness Landscape Analysis of a tRNA Gene Reveals that the Wild Type Allele is Sub-optimal, Yet Mutationally Robust

Tzahi Gabzi,[1,2] Yitzhak Pilpel,*[,1] and Tamar Friedlander *[,2]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 7610001, Israel

[2]The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, Faculty of Agriculture, Hebrew University of Jerusalem, 229 Herzl St., Rehovot 7610001, Israel

***Corresponding authors:** E-mails: tamar.friedlander@mail.huji.ac.il; pilpel@weizmann.ac.il.

**Associate editor:** Deepa Agashe

## Abstract

**Fitness landscape mapping and the prediction of evolutionary trajectories on these landscapes are major tasks in evolutionary biology research. Evolutionary dynamics is tightly linked to the landscape topography, but this relation is not straightforward. Here, we analyze a fitness landscape of a yeast tRNA gene, previously measured under four different conditions. We find that the wild type allele is sub-optimal, and 8–10% of its variants are fitter. We rule out the possibilities that the wild type is fittest on average on these four conditions or located on a local fitness maximum. Notwithstanding, we cannot exclude the possibility that the wild type might be fittest in some of the many conditions in the complex ecology that yeast lives at. Instead, we find that the wild type is mutationally robust ("flat"), while more fit variants are typically mutationally fragile. Similar observations of mutational robustness or flatness have been so far made in very few cases, predominantly in viral genomes.**

***Key words:*** **computational biology, population genetics, molecular evolution, fitness landscapes.**

## Introduction

Fitness landscape mapping and prediction of evolutionary trajectories on these landscapes are major tasks in evolutionary biology (Wright 1932). While evolutionary theory predicts that population mean fitness should increase over time, it offers only few quantitative predictions for the dynamics of evolution and the possible evolutionary trajectories. The main hurdle for generally computing evolutionary trajectories is their dependence on the underlying fitness landscape. Currently available fitness landscapes include between 16 and 100,000 different genotypes (for review, see Szendro et al. 2013; de Visser and Krug 2014; Obolski et al. 2018). Yet, even the largest datasets (Jacquier et al. 2013; Roscoe et al. 2013; Puchta et al. 2016; Sarkisyan et al. 2016) encompass only small fractions of the entire fitness landscape of even a single gene. As detailed fitness measurements have been unavailable until recently, most of the associated theory was developed in isolation from data (Kingman 1978; Kauffman and Levin 1987; Kauffman and Weinberger 1989; Weinreich 2005; Park and Krug 2008; Kryazhimskiy et al. 2009; Weissman et al. 2009; McCandlish 2013, 2018).

The advent of sequencing technologies now enables measurement of increasingly larger fitness landscape datasets (Puchta et al. 2016; Sarkisyan et al. 2016; Somermeyer et al. 2022). It is then desirable to predict evolutionary trajectories on these empirical fitness landscapes, using the previously developed theory in this field.

A recent set of experiments characterized the fitness landscape of the $tRNA^{Arg}_{CCU}$ gene of *Saccharomyces cerevisiae*. As this gene is relatively short (72 nucleotides), its landscape is significantly smaller than that of a typical protein. It is a single-copy, non-essential gene, such that many of its mutants are viable. Li et al. measured the growth rates of 23,284 different variants of this gene (fig. 1A) under four different growth conditions (23 °C, 30 °C, 37 °C and oxidative stress) (Li et al. 2016; Li and Zhang 2018). The richness of this dataset renders it a highly valuable case study for analyzing topographic properties and evolutionary trajectories of an empirical fitness landscape and for comparing them with theoretical predictions. In analyzing this fitness landscape, we noticed that many variants appear fitter than the wild type in each of the examined conditions. The wild type's advantage appears instead to be in its mutational robustness, since its neighbors in sequence space are relatively fit, too.

## Results

In analysis of the tRNA fitness landscape across four different growth conditions (Li et al. 2016; Li and Zhang 2018), we made a remarkable observation that the wild type is not the genotype with highest fitness under any of the four conditions. Under each of the conditions, between 8% and 10% of the variants exhibited higher fitness than the wild type (fig. 1B). We then analyzed possible sources





Letter

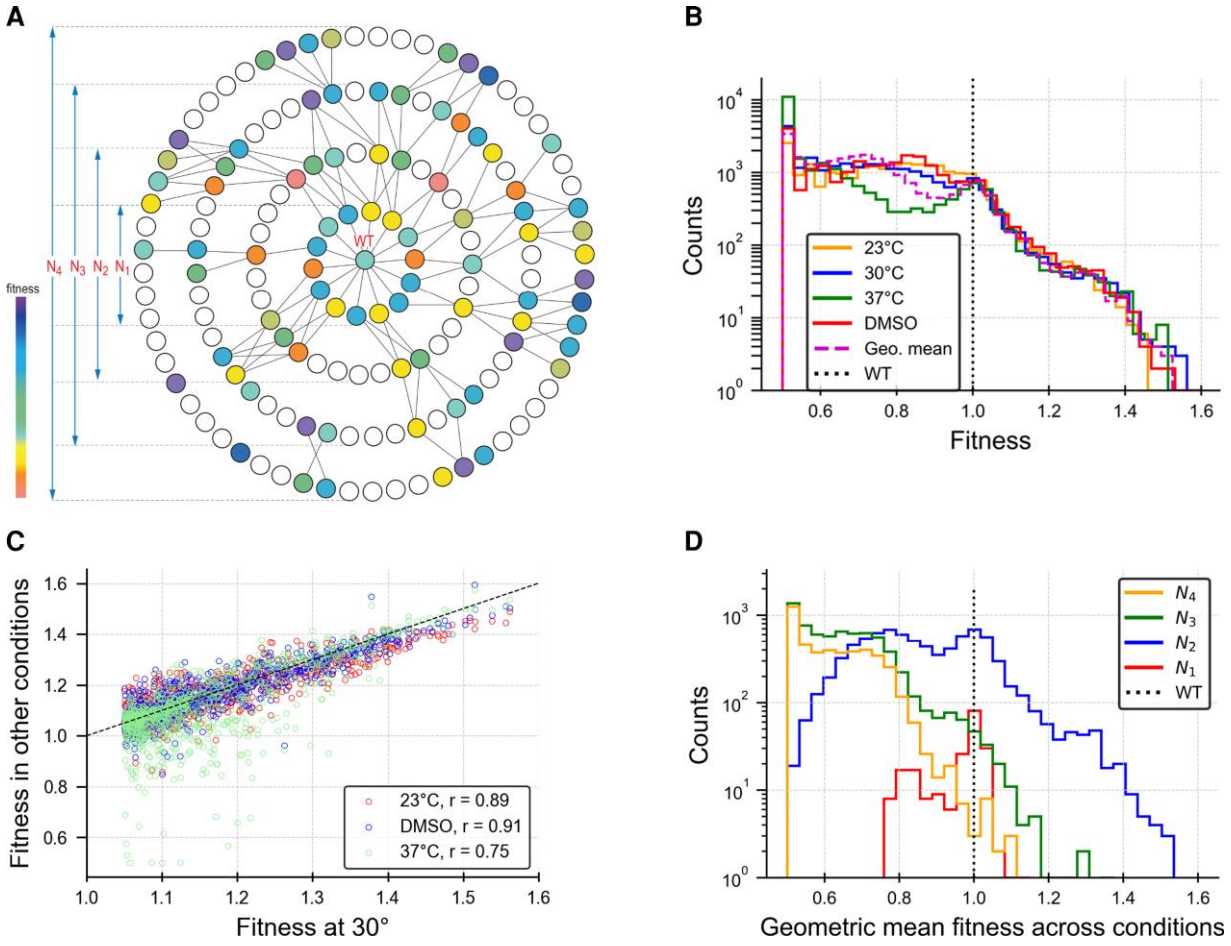


**Fig. 1.** The wild type is not the fittest under any of the conditions or on average on all four. (*A*) A schematic visualization of the experimentally measured tRNA fitness landscape. Each circle represents a genotype. Filled circles represent genotypes whose fitness values (here encoded by different colors) were measured. Empty circles represent genotypes whose fitness values were not measured. We use here a concentric representation of the fitness landscape, centered around the wild type, where the minimal number of steps on the graph between any two genotypes is the number of point mutations separating them. The wild type is then surrounded by expanding circles of its single mutants (denoted by $N_1$), double mutants ($N_2$), etc. The experiment probed all the wild type's single-point mutants, but only decreasingly smaller proportions of the following mutational neighborhoods, $N_i$. (*B*) The distributions of all fitness values measured under four different conditions (23 °C, 30 °C, 37 °C and DMSO), and the distribution of $\langle f_i \rangle$, their geometric mean. The wild type fitness value is shown by the dotted vertical line. Fitness was defined relative to the wild type's fitness, such that the wild type fitness was set to 1 for each condition. Under each of the conditions tested, 8–10% of the genotypes in this dataset were fitter than the wild type. (*C*) Fitness values of variants with fitness in the range [1.05, 1.6] at 30 °C plotted against their fitness values at 23 °C, DMSO and 37 °C. The correlation coefficients between fitness values under different conditions were $r = 0.89, 0.91, 0.75$ respectively. The $x = y$ line is shown for reference (black dashed line). We conclude that variants that have high fitness in one condition usually have high fitness in all four of them. (*D*) Geometric mean fitness value histograms of genotypes in the wild type's four mutational neighborhoods $N_1 - N_4$. The wild type fitness (dotted vertical line) is shown for reference. Notably, all four mutational neighborhoods contain fitter than wild type genotypes, but the largest proportion of fitter variants is in $N_2$.

for measurement errors, including statistical sampling fluctuations in read-counts, as a source of inaccuracy in fitness assessment and the possibility that the fitness effect was due to independent mutations that fortuitously occurred elsewhere in the genome (supplementary text, Supplementary Material online). We conclude that although such errors do exist, they could not fully account for the wild type's fitness sub-optimality.

A possible explanation for the apparent sub-optimality of the wild type could be that while some variants are fitter than the wild type under a specific condition, they are much less fit under other conditions, such that, *on average* across conditions the wild type is fittest. To test this explanation, we checked for all high-fitness genotypes ($f > 1.05$ at 30 °C) the correlation between their fitness values under the various growth conditions—figure 1C. We found, that most genotypes which are fit under one condition are also fit under others.

To formally compare between fitness values averaged over multiple conditions, we also calculated the geometric mean fitness (Gillespie 2004), $\langle f_i \rangle = (\prod_m f_i^m)^{1/M}$, where $f_i^m$ is the fitness value of the $i$th genotype in the $m$th condition out of $M$ (see supplementary text, Supplementary Material online).

Figure 1*B* also shows a histogram of the geometric mean fitness values $\langle f_i \rangle$ of all the genotypes in our dataset. Here

too, we observe that the wild type is not the fittest across conditions, but 2098 variants (9%) have higher geometric mean fitness than the wild type's.

Both results argue against the possibility that the wild type is the fittest on average across conditions. A possible caveat is that only four conditions were included in this calculation which might not fully represent yeast natural habitat. For instance, some of the high-fitness variants could be inferior in another condition not included in this experiment.

Alternatively, the wild type sub-optimality could hypothetically be rooted in the fitness landscape topography. If, for example, the wild type were an isolated local maximum, separated from the global fitness maximum by fitness valleys, the population could be "trapped" in the current wild type genotype, hindered from reaching the global maximum (at least temporarily) (Kauffman and Levin 1987; Weissman et al. 2009). To test this hypothesis, we started by locating the high-fitness genotypes in the dataset. We define "Mutational neighborhoods" $N_i(WT)$ surrounding the wild type as the set of genotypes reachable by $i$ point mutations (shortest path) from the wild type (see fig. 1A). Figure 1D shows the fitness distributions of the four mutational neighborhoods $N_1 - N_4$ (single to quadruple mutants). We found that all four mutational neighborhoods contained fitter-than wild type genotypes, but the largest proportion of such fitter genotypes was in $N_2$, only two point mutations away from the wild type.

Dissection of each mutational neighborhood into one of three fitness categories shows that 51% (106 out of 207) of the wild type's single mutants have similar geometric mean fitness values to the wild type's ($|\langle f_i\rangle - \langle f(WT)\rangle| < \Delta^{(1)}$), with $\Delta^{(1)} = 0.023$, 37% (77 genotypes) of them were much less fit ($\langle f_i\rangle \leq \langle f(WT)\rangle - \Delta^{(1)}$), and 12% (24 genotypes) were fitter than the wild type by more than $\Delta^{(1)} = 0.023$, ($\langle f_i\rangle > \langle f(WT)\rangle + \Delta^{(1)}$). Amongst the $N_2(WT)$ genotypes (wild type's double mutants) the proportion of such fitter-than wild type genotypes ($\langle f_i\rangle > \langle f(WT)\rangle + \Delta^{(1)}$) was even larger (1,395 out of 8,101; 17%) and even amongst $N_3(WT)$ (triple-mutants) we found 1% (71 out of 6,891) having fitness values higher than the wild type's by at least $\Delta^{(1)}$. The value of $\Delta^{(1)} = 0.023$ is equivalent to 0.95 confidence that the geometric mean fitness of one genotype is larger than that of the other (see supplementary text, Supplementary Material online).

We then checked for the existence of evolutionary trajectories of non-decreasing fitness, leading from the wild type to the fitter genotypes in $N_2$ and $N_3$. We mapped all 2- and 3-step trajectories of strictly increasing geometric mean fitness, originating from the wild type. The number of fitness increasing trajectories depends on the minimal required fitness difference between consecutive genotypes in the trajectory. By requiring a minimal fitness increase $\Delta$, we ensure with certain confidence level, that such trajectories are strictly increasing, and are not mistakenly classified as such due to inaccuracies in any of the fitness estimations along the trajectory. To obtain statistical significance of 0.95 for the whole trajectory, $\Delta^{(2)} = 0.028$ is required for each step in a 2-step trajectory and $\Delta^{(3)} = 0.0306$ for each step in a 3-step trajectory (see supplementary text, Supplementary Material online). We find 142 different 2-step trajectories and 17 different 3-step trajectories meeting the 0.95 confidence criterion. This requirement of $\Delta$ is very conservative, and its relaxation (smaller $\Delta$) significantly augments the trajectory count. Our trajectory count is certainly an underestimation of the number of fitness-increasing trajectories for two reasons. Firstly, here we only accounted for trajectories that are fully included in our incomplete dataset, whereas expansion of the landscape, by measuring the fitness of additional variants, is expected to have increased the count of such fitness increasing paths. Secondly, neutral evolutionary transitions to genotypes with equal or nearly-equal fitness are also possible and were not included in this enumeration.

Hence, we conclude that the wild type is not a local maximum which is mutationally isolated from higher-fitness genotypes. It is worth mentioning in this context that the higher the landscape dimension is, the larger the number of possible single mutants for each genotype is. A genotype is only a local maximum if *all* its single mutants are less fit. Hence, with the increase of landscape dimensionality, it is less likely to find local maxima (Obolski et al. 2018).

Up to this point, we found that the wild type is neither the fittest on average, nor is it a local maximum. What is then unique about this genotype and why was this sequence selected in evolution to be the wild type? In figure 1D, we saw already that the fitness distribution of the wild type single mutants $N_1$ is narrower compared with the fitness value distributions of further mutants ($N_2 - N_4$).

Hence, we next sought to characterize whether such mutational robustness is common in this fitness landscape or whether the wild type is unique in residing in a relatively flat region of the landscape. We defined genotype mutational fragility as the average fitness difference between the genotype and its deleterious single-mutants (beneficial single-mutants receive zero weight):

$$\phi_i = \frac{1}{|N_1(i)|}\sum_{j\in N_1(i)} |f_i - f_j|_+, \qquad |x|_+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{1}$$

where $|N_1(i)|$ is the total number of single mutants of genotype $i$. What is the significance of this new measure? To demonstrate its behavior and relation to fitness, we began by applying it on a simulated NK model (Kauffman and Weinberger 1989) landscape (fig. 2A and B). We observe a general trend where fragility increases with fitness. For low-fitness genotypes fragility is nearly zero, because their single mutants are mostly fitter and hence disregarded in this measure. Conversely, fragility is highest for high-fitness genotypes, because their single-mutants are

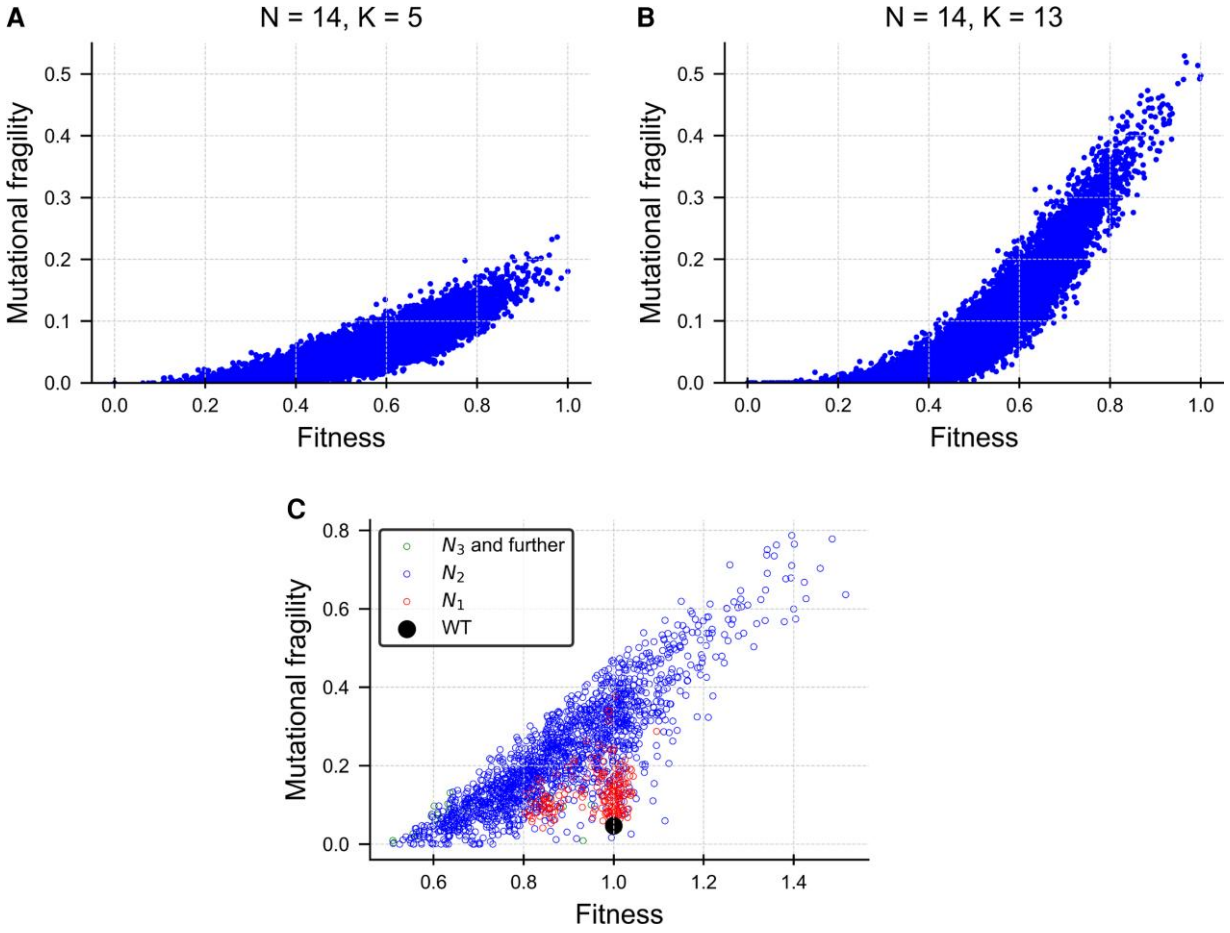


FIG. 2. Mutational fragility. Fragility vs. fitness scatter plots (*A*,*B*) in a simulated NK landscape with parameter values $N = 14$, $K = 5$ (*A*) and $K = 13$ (*B*). Low-fitness genotypes have mostly beneficial single-mutants and hence low mutational fragility (zero if all mutants are beneficial). The higher the fitness of a genotype, the more deleterious its single-mutants are, and hence it is more fragile. This effect is more pronounced in the uncorrelated landscape ($K = N - 1 = 13$, (*B*)) but is still observed in the partially correlated landscape with $K = 5$ (*A*). (*C*) Genotype mutational fragility against fitness in the tRNA dataset – scatter plot. The wild type (black) is nearly the least fragile to deleterious mutations amongst all genotypes with similar fitness values. We show here all genotypes with at least 5 single-mutants whose fitness is included in the dataset, such that at least 3 of them are further away from the wild type (see Methods). We illustrate genotypes that are wild type single, double, or higher mutants using different colors.

mostly less fit. The parameter $K$ is used to tune the correlation between fitness values of adjacent genotypes (namely, genotypes that differ in a small number of positions). A $K = 5$ (fig. 2*A*) landscape exhibits partial correlation, and can hence be regarded as "smooth," while the $K = 13$ landscape is "rugged" in the sense that fitness values of neighboring sequences are uncorrelated (fig. 2*B*) (see Methods for definition). As expected, the differences in fragility between the intermediate and the extreme fitness genotypes were larger in the uncorrelated landscape (fig. 2*B*) than in the partially-correlated one (fig. 2*A*), while their fitness value distributions were similar (supplementary fig. S5, Supplementary Material online). For most fitness values, we find a range of fragility values. In both landscapes we observe, that the highest fitness value for which genotypes with zero fragility still exist are the intermediate ones ($f \approx 0.5$ here).

Returning to the tRNA landscape, the fragility definition (1) is based on fitness information of all the single mutants of genotype $i$. In practice, with the exception of the wild type, we only had measurements of a subset of the single mutants and estimated $\phi_i$ using partial data. To minimize biases because of small numbers of single mutants, fragility was only calculated for genotypes having at least five single-mutants, three of which are further away from the wild type (Methods). This limitation enabled calculation only for 1,854 variants out of 23,284.

Figure 2C shows a scatter plot of genotype mutational fragility plotted against genotype fitness. Interestingly, we observe that the wild type is at the "tip" of the fragility-fitness cloud, such that it is nearly the least fragile amongst genotypes with similar fitness value and nearly the fittest amongst genotypes with similarly low fragility. Hence, it exhibits a balance between fitness and mutational robustness.

## Discussion

Recent advances in high-throughput experimental methods have allowed for large-scale characterization of

empirical fitness landscapes (Jacquier et al. 2013; Roscoe et al. 2013; Puchta et al. 2016; Sarkisyan et al. 2016), which can be applied to test hypotheses about the driving forces of evolutionary dynamics. Here we found a wild type which is sub-optimal and ruled out the possibilities that it is the fittest on average on multiple conditions or located on a local fitness maximum. Instead, we found that the wild type is amongst the least mutationally fragile genotypes in the dataset, balancing between fitness and mutational robustness, but the evolutionary mechanism at play here remains elusive. Robust (flat) genotypes were predicted theoretically (Swetina and Schuster 1982; Sardanyés et al. 2008; Beardmore et al. 2011), but were previously reported only in organisms having high mutation rates such as viruses (Codoñer et al. 2006; Sanjuán et al. 2007) or in digital organisms (Wilke et al. 2001). To the best of our knowledge, this is the first report of a flat gene in a low-mutation rate organism. It is increasingly appreciated that mutation rates are non-uniform across the genome and specifically, tRNA genes were shown to have 7–10-fold higher mutation rates compared with the background genome (Saini et al. 2017; Thornlow et al. 2018). Yet, it is unclear whether such a mutation rate is sufficient to select for flatness (Swetina and Schuster 1982). One possibility that we cannot exclude, is that the wild type could be fittest in some of the many conditions in the complex ecology that yeast lives at (Liti 2015), that might not be captured by the experimental choice of conditions. An additional explanation, which does not rely on high-mutation rate, is that RNA genes are selected not only for their fitness, but also for the molecule thermodynamic stability. Evidence exists that thermodynamic stability of RNAs, may correlate also with their evolutionary stability, namely their robustness to mutations (Ancel et al. 2000; Meyers et al. 2004). Although the source of this correlation is not entirely understood, we mention the possibility that since RNAs, and tRNA included, need to maintain thermodynamic stability, this constraint may have also shaped them as evolutionarily stable and mutationally non-fragile.

As the number of large-scale fitness measurements of particular landscapes is still limited, additional examples for wild type alleles being sub-optimal are scarce, but see Bank et al. (2016) and Bershtein et al. (2015).

Inherent to the astronomical dimensionality of fitness landscapes is our inability to fully measure them (de Visser and Krug 2014). Full landscape mappings are possible only for computationally fabricated landscapes (Kauffman and Levin 1987; Friedlander et al. 2017; Collins-Hed and Ardell 2019). Thus, usage of fragments of a fitness landscape to draw general conclusions is a common practice in the field. It does raise the fundamental question whether indeed it represents the entirety of the landscape and hence, should be used with caution. Recent works handled the sparse sampling of fitness datasets by sampling around multiple focal genotypes (Somermeyer et al. 2022) or by interpolating between the measured genotypes to estimate fitness values of missing ones (Hopf et al. 2017; Zhou and McCandlish 2020) with some success. While the latter techniques are computationally very demanding, it would be interesting to test in the future whether they are applicable for computing evolutionary dynamics on incomplete landscapes.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution online*.

## Acknowledgments

We thank Chuan Li and Jianzhi Zhang for sharing their experimental data, Ohad-Noy Feldheim for help with the fragility calculation and Yuval Benjamini for consultation regarding statistical analysis. We thank Yoav Ram and Daniel Weissman for comments on an earlier version of the manuscript. We thank and Johannes Berg and Peter Stadler for bringing to our attention the literature on RNA thermodynamic stability. T.F. acknowledges funding from the Israel Science Foundation (grant no. 1889/19) and from the Hebrew University of Jerusalem. Y.P. acknowledges support from the Minerva Foundation. Y.P. is a Kimmel Investigator and an incumbent of the Ben May Professorial Chair.

## Data Availability

The code we used for data analysis and creation of the figures can be found in https://github.com/Tamar-Friedlander-Lab/tRNA-fitness-landscape-MBE.

## References

Ancel LW, Fontana W. 2000. Plasticity, evolvability, and modularity in RNA. *J Exp Zool*. **288**(3):242–283.

Bank C, Matuszewski S, Hietpas RT, Jensen JD. 2016. On the (un)predictability of a large intragenic fitness landscape. *Proc Natl Acad Sci*. **113**(49):14085–14090.

Beardmore RE, Gudelj I, Lipson DA, Hurst LD. 2011. Metabolic trade-offs and the maintenance of the fittest and the flattest. *Nature* **472**(7343):342–346.

Bershtein S, Serohijos AWR, Bhattacharyya S, Manhart M, Choi J-M, Mu W, Zhou J, Shakhnovich EI. 2015. Protein homeostasis imposes a barrier on functional integration of horizontally transferred genes in bacteria. *PLOS Genet*. **11**(10):e1005612.

Codoñer FM, Darós J-A, Solé RV, Elena SF. 2006. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLOS Pathog*. **2**(12):e136.

Collins-Hed AI, Ardell DH. 2019. Match fitness landscapes for macromolecular interaction networks: Selection for translational accuracy and rate can displace tRNA-binding interfaces of non-cognate aminoacyl-tRNA synthetases. *Theor Popul Biol*. **129**:68–80.

Friedlander T, Prizak R, Barton NH, Tkačik G. 2017. Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nat Commun*. **8**(1):216.

Gillespie JH. 2004. *Population genetics: a concise guide*. 2nd ed. Baltimore (MD): The Johns Hopkins University Press.

Hopf TA, Ingraham JB, Poelwijk FJ, Schärfe CPI, Springer M, Sander C, Marks DS. 2017. Mutation effects predicted from sequence covariation. *Nat Biotechnol*. **35**(2):128–135.

Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, Bercot B, Petit E, Poulain J, Barnaud G, *et al.* 2013. Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci.* **110**(32):13067–13072.

Kauffman S, Levin S. 1987. Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol.* **128**(1):11–45.

Kauffman SA, Weinberger ED. 1989. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J Theor Biol.* **141**(2):211–245.

Kingman JFC. 1978. A simple model for the balance between selection and mutation. *J Appl Probab.* **15**(1):1–12.

Kryazhimskiy S, Tkačik G, Plotkin JB. 2009. The dynamics of adaptation on correlated fitness landscapes. *Proc Natl Acad Sci.* **106**(44):18638–18643.

Li C, Qian W, Maclean CJ, Zhang J. 2016. The fitness landscape of a tRNA gene. *Science* **352**(6287):837–840.

Li C, Zhang J. 2018. Multi-environment fitness landscapes of a tRNA gene. *Nat Ecol Evol.* **2**(6):1025.

Liti G. 2015. The fascinating and secret wild life of the budding yeast *S. cerevisiae*. *eLife* **4**:e05835.

McCandlish DM. 2013. On the findability of genotypes. *Evolution* **67**(9):2592–2603.

McCandlish DM. 2018. Long-term evolution on complex fitness landscapes when mutation is weak. *Heredity* **121**(5):449–465.

Meyers LA, Lee JF, Cowperthwaite M, Ellington AD. 2004. The robustness of naturally and artificially selected nucleic acid secondary structures. *J Mol Evol.* **58**(6):681–691.

Obolski U, Ram Y, Hadany L. 2018. Key issues review: evolution on rugged adaptive landscapes. *Rep Prog Phys.* **81**(1):012602.

Park S-C, Krug J. 2008. Evolution in random fitness landscapes: the infinite sites model. *J Stat Mech: Theory Exp.* **2008**(04):P04014.

Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016. Network of epistatic interactions within a yeast snoRNA. *Science* **352**(6287):840–844.

Roscoe BP, Thayer KM, Zeldovich KB, Fushman D, Bolon DNA. 2013. Analyses of the effects of all ubiquitin point mutants on yeast growth rate. *J Mol Biol.* **425**(8):1363–1377.

Saini N, Roberts SA, Sterling JF, Malc EP, Mieczkowski PA, Gordenin DA. 2017. APOBEC3B cytidine deaminase targets the non-transcribed strand of tRNA genes in yeast. *DNA Repair* **53**:4–14.

Sanjuán R, Cuevas JM, Furió V, Holmes EC, Moya A. 2007. Selection for robustness in mutagenized RNA viruses. *PLOS Genet.* **3**(6):e93.

Sardanyés J, Elena SF, Solé RV. 2008. Simple quasispecies models for the survival-of-the-flattest effect: the role of space. *J Theor Biol.* **250**(3):560–568.

Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, *et al.* 2016. Local fitness landscape of the green fluorescent protein. *Nature* **533**(7603):397–401.

Somermeyer LG, Fleiss A, Mishin AS, Bozhanova NG, Igolkina AA, Meiler J, Pujol M-EA, Putintseva EV, Sarkisyan KS, Kondrashov FA. 2022. Heterogeneity of the GFP fitness landscape and data-driven protein design. *eLife* **11**:e75842.

Swetina J, Schuster P. 1982. Self-replication with errors: a model for polvnucleotide replication. This paper is considered as part II of model Studies on RNA replication. Part I is the Gassner and Schuster [14]. *Biophys Chem.* **16**(4):329–345.

Szendro IG, Schenk MF, Franke J, Krug J, de Visser JGM. 2013. Quantitative analyses of empirical fitness landscapes. *J Stat Mech: Theory Exp.* **2013**(01):P01005.

Thornlow BP, Hough J, Roger JM, Gong H, Lowe TM, Corbett-Detig RB. 2018. Transfer RNA genes experience exceptionally elevated mutation rates. *Proc Natl Acad Sci.* **115-**(36):8996–9001.

de Visser JGM, Krug J. 2014. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet.* **15**(7):480–490.

Weinreich DM. 2005. The rank ordering of genotypic fitness values predicts genetic constraint on natural selection on landscapes lacking sign epistasis. *Genetics* **171**(3):1397–1405.

Weissman DB, Desai MM, Fisher DS, Feldman MW. 2009. The rate at which asexual populations cross fitness valleys. *Theor Popul Biol.* **75**(4):286–300.

Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C. 2001. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**(6844):331–333.

Wright S. 1932. The roles of mutation, inbreeding,crossbreeding and selection in evolution. Proc. 6th Int. Congress Genet. **1**:356–366.

Zhou J, McCandlish DM. 2020. Minimum epistasis interpolation for sequence-function relationships. *Nat Commun.* **11**(1):1782.