



מכון ויצמן למדע

WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree
Master of Science

עבודת גמר (תזה) לתואר
מוסמך למדעים

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Tzahi Gabzi

מאת
צחי גבזי

חקר של 'פונקציות כשירות' לא חלקות בעל היכולת להסביר חוסר
אופטימליות של זן בר

Exploration of Rugged Fitness Landscapes Explains Wild Type's Sub Optimality

Advisors:
Prof. Yitzhak Pilpel
Dr. Tamar Friedlander

מנחים:
פרופ' צחי פלפל
ד"ר תמר פרידלנדר

February 2019

אדר התשע"ט

Exploration of Rugged Fitness Landscapes Reveals Wild Type's Sub Optimality

March 18, 2019

Abstract

Evolution is a powerful engine that optimizes biological solutions to environmental challenges. "Survival of the Fittest" implies that the most fit individual will always be selected for. Yet, along with selection that would favor the fittest, mutations may erode it. Thus, the balance between selection and mutation might result in survival of the less-than-fittest at certain circumstances. In particular, the quasi-species model that presents a mathematical description of the mutation-selection balance predicts that under certain circumstances, the fittest might not be selected for but rather the "flattest" - a gene or a genome, which has sub-optimal fitness, yet is surrounded in the space of sequences by other sequence neighbors that are similarly sub-optimal. The "Survival of the Flattest" theory was suggested years ago but was so far not examined with experiments on living organisms.

Today, synthetic biology allows fabricating numerous sequence variants of a gene and measuring the fitness of each in living cells. In particular, Chuan et al. recently mapped the fitness landscape (FLS) of a short yeast gene - a $tRNA^{CCU-Arg}$ in four environmental conditions. Interestingly, in all conditions, the Wild Type (WT) was not the fittest variant found. Motivated by their experimental results, we computationally explored evolutionary dynamics on this FLS and characterized its structural properties. We also compared this FLS with theoretical ones, studied in the literature, such as the NK model.

We defined a local measure of genotype and environmental FLS flatness as the averaged gradients with respect to nearest genotype neighbors (fixed environment) and environment neighbors (fixed genotype) respectively and found that the WT is relatively flat. i.e. robust to mutational and environmental changes. Simply put, the WT fitness will change relatively slightly if a mutation occurs or the environment changes. We then searched, using various simulations, for proper settings in which these properties are evolutionary advantageous to find it is probably so for $tRNA^{CCU-Arg}$ in high mutation rate.

Contents

1	Introduction	4
1.1	Definitions and Notations	5
1.2	Evolutionary Dynamics on Fitness Landscapes, The Quasi Species Theory	7
1.3	Chuan et al. tRNA FLS Study	9
2	Goals	11
3	Results - Characterization of Chuan et al. Data	12
3.1	The WT is not the Fittest	12
3.2	The Validity of the Experimental Observation	12
3.2.1	Background Mutations	14
3.2.2	Read-Count Sampling Noise Effect	16
3.3	The 'Anna Karenina Effect' and Data Sparseness	19
4	Results - Looking for Wild Type's Evolutionary Advantages	24
4.1	Wild Type is (not) the fittest across conditions	24
4.2	Evolution (didn't) 'get stuck'	26
4.3	The WT is robust to mutations (Genotype Flatness) in each condition	27
4.3.1	Dealing with sampling bias	28
4.4	WT is robust to environmental changes (Environment Flatness)	32
4.5	Does robustness grant evolutionary advantage?	33
4.5.1	Simulation Results	34
4.5.2	Bridging Mutation Gap	35
4.5.3	Is High Connectivity the Culprit?	37
4.5.4	NK Model Simulations, Connectivity and Flatness Decoupling	39
5	Discussion	43
6	Materials and Methods	45
6.1	tRNA Fitness Values Alignment Across Conditions	45
6.2	QS simulation on the tRNA FLS	48
7	Acknowledgments	54
A	Fitness Landscapes as an Optimization Problem	55
B	snoRNA and GFP FLSs	56
C	Fitness correlations between different repetitions	62

Nomenclature

\bar{f} Average population fitness

A Alphabet set

f_i Fitness of genotype i

h_{ji} Hamming distance between genotypes j and i

L Genome Length

N_k The k neighborhood of a genotype

Q Quality coefficient - the probability of a genotype to replicate with no error

q Mutation rate per base pair

q_c Critical mutation rate above which Survival of the Flattest effect is in force

q_e Mutation error threshold above which population collapses

q_{ji} Transition probability (mutation rate) from genotype j to i

ABM Agent Based Modeling

FLS Fitness Landscape

HOC House of Cards (FLS Model)

MFT Mean Field Theory

nt nucleotide

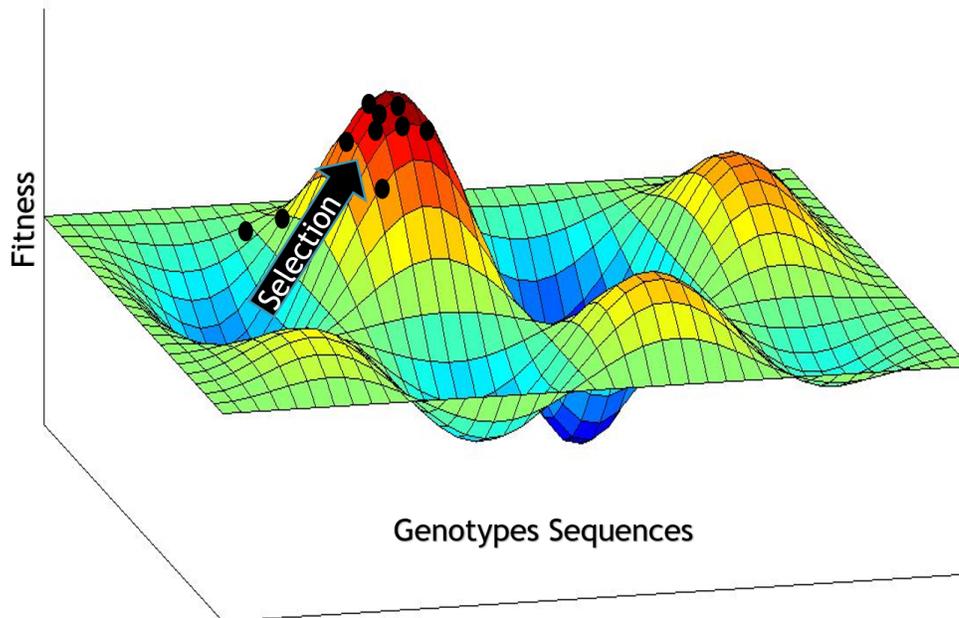
QS Quasi Species (Theory)

RMF Rough Mount Fuji (FLS Model)

SS Steady State

WT Wild type

Figure 1:



Schematic Sketch of evolutionary dynamics on a Fitness Landscape. This function depicts how the fitness of different variants of a given genome changes. Natural selection drives populations to higher fitness values as stated in the fundamental theorem of natural selection (see for example [8], chapter 3). Here we assume that the genome alone fully determines the fitness, neglecting non genetic effects such as the environment.

1 Introduction

Since Darwin's 'Origin of species', many researchers have tried in various ways to model the evolutionary process. One of the first and prominent ways was suggested back in the 1930's by Sewall Wright - the 'Adaptive Landscape' (or 'Fitness Landscape') [40]. By fixing the environmental conditions, we can assign each of the possible genetic sequences of an organism some real scalar, representing fitness value. Here fitness is defined as the reproductive rate of the organism, when we assume that the genome alone determines fitness values and that they are fixed. This function defines the Fitness Landscape (**FLS**, See Fig. 1). We can then view the evolutionary process of mutation and selection as a stochastic optimization process that searches for the highest point(s) on the FLS. For example, imagine a population of organisms born with similar genetic sequences to a very hot environment. It might be that most of the population will perish due to low fitness, and only the lucky rare few with a desired mutation will survive and give birth to an adapted generation. they are the fittest. In this manner, over time, the population will increase its fitness and 'climb up' the FLS to one of its peaks.

The FLS metaphor, since its inception, first gave rise to theoretical models trying to capture its structure, hoping to gain new insights into evolutionary dynamics and even predict it to some extent. These models range from a naive approach in which fitness values drawn uniformly at random are assigned to binary sequences (playing the role of genes) as the House Of Cards model (HOC) to biologically motivated approaches such as the NK model (See also page 55) or Rough Mount Fuji (RMF) [11, 12, 21, 7, 6, 14, 9].

Little was known about **real** FLSs up until recently, when technological advancements made FLS studies of unprecedented scale possible. Today, synthetic biology allows fabricating tens of thousands (as opposed to dozens or hundreds, not so long ago) sequence variants of a gene and measuring the fitness of each in living cells [31, 27, 17, 18]. This, in turn, encouraged further theoretical studies that tried to calculate topographic properties of experimental and artificial FLSs in order to classify them in various ways [34, 35, 22, 38]. The large amount of data motivated researches to obtain deeper insights into (asexual) evolutionary dynamics underlying major biological process, result in a plethora of works on the subject [24, 4, 5].

In particular, Chuan et al. recently mapped the local FLS of a non-essential short 72 nucleotide long yeast *tRNA^{CCU-Arg}* gene (See Fig. 9), in four environmental conditions [17, 18]. In their studies, they reported significant negatively biased epistasis of almost half of all mutation pairs at 37C, and revealed a simple genotype-by-environment interaction which enabled them to transform FLS in one condition to another. This work continue on to explore their FLS structure and dynamics upon it. However, before we go on to thoroughly present their experiments and their advantages to our needs (See page 9), a more formal framework and definitions are needed.

1.1 Definitions and Notations

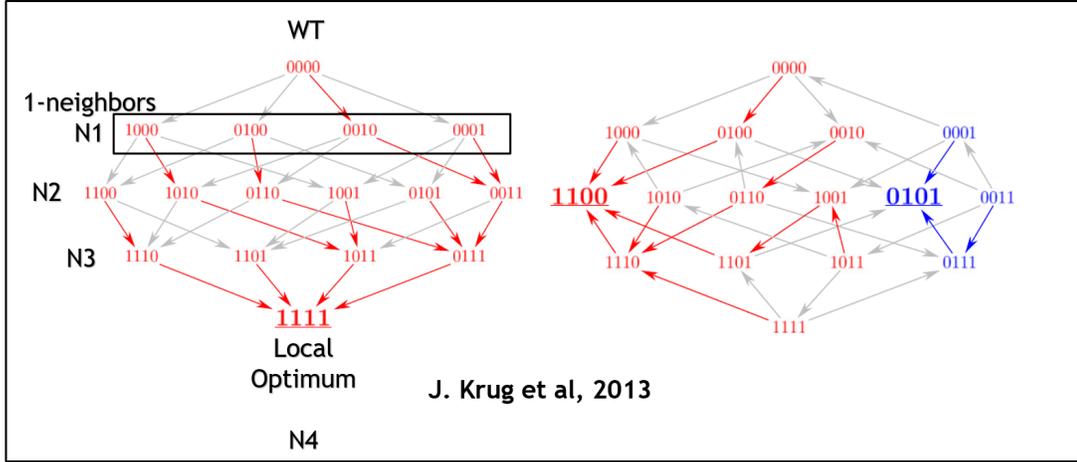
Fitness Landscapes are multidimensional entities which are hard to grasp and visualize. Thus a dilemma arise: What's the best way to mediate concrete definitions and notations to the reader? One way to go about it is the formal approach, using strict mathematical definitions. It is more rigorous, leaves no place for doubt, but more cumbersome for audience not familiar with it (Still, for a formal mathematical treatment see Appendix A on page 55). Another way, which we found more attractive, is using an example of a simple Fitness Landscape to assist us explain and illustrate important terms that we will later use.

Fig. 2 presents Fitness Landscapes of *Methylobacterium extorquens* and Malaria drug resistance gene adapted from Szendro et al. [36]. These empirical FLSs contain all combinations of mutations at 4 loci. Genotypes are represented by binary sequences, where 0 (1) indicates the absence (presence) of the corresponding mutation. Hence there are $2^4 = 16$ sequences presented in each panel. Arrows point in the direction of increasing fitness.

The original non-mutated sequence (0000 in both cases) will be referred to as the **Wild Type (WT)**. All 16 genotypes (the WT and its 15 possible sequence mutants) are part of what we call the **sequence domain**. The sequence domain is the set of all possible sequences, joined together with a way to measure distance between them. In this particular Fitness Landscape we measure distance using Hamming distance which just counts the minimal number of positions (sites) needed to be changed to transform one sequence to become the other. For instance, the distance between 0000 (WT) to 1001 is 2. By assigning a fitness value to each of the sequences, we obtain a **Fitness Landscape**.

Moreover, two sequences will be called ***k*-neighbors** if they are at a (Hamming) distance *k* from each other. The *k*-neighborhood of a genotype, which we will denote by N_k , is the set of all its *k*-neighbors. Underlined nodes in Fig. 2 correspond to fitness **local optima** which are genotypes with higher fitness than all of their N_1 neighbors. Colored arrows point towards the fittest neighbor.

Figure 2:



Simple Experimentally Measured Fitness Landscape. *Methylobacterium extorquens* (left) and Malaria drug resistance gene (right) FLSs are presented, adopted from Szendro et al. [36]. These empirical FLSs contain all combinations of mutations at 4 loci. Genotypes are represented by binary sequences, where 0 (1) indicates the absence (presence) of the corresponding mutation. Arrows point in the direction of increasing fitness, colored arrows point towards the fittest neighbor. Underlined nodes correspond to fitness local optima.

In this particular example, we can use binary sequences to denote mutations to the WT since there are only two options for each position: mutation / no mutation. Of course, other options are also possible, depending on the type of sequence. To symbolize nucleotide sequences and their point mutations (assuming no indels) we can use a 4-letter alphabet and for amino-acid sequences a 20-letter one. However, by working with larger alphabet sizes we obtain an much larger sequence space. This is naturally the case (exponentially larger sequence space) when working also with longer sequence size. Generally, the size of the sequence domain is given by $|A|^L$ when $|A|$ is the alphabet size and L is the sequence length. The size of k -neighborhood N_k is then:

$$|N_k| = (|A| - 1)^k \binom{L}{k}. \quad (1)$$

For example, even a short 100 nucleotide (nt) gene gives rise to a staggering $4^{100} \approx 10^{30}$ possible sequences, demonstrating the difficulty of studying and grasping real FLSs. Szendro's example is a small fragment of FLS in which $L = 4, |A| = 2$, which also makes it easy to measure and visualize. However when L or $|A|$ are large, measurements can only map a small fraction of the actual FLS. Only local measurements, which are densely located around a few chosen genotypes, or a relatively small number of random measurements from the whole FLS are possible. The lack of complete FLS measurements (a.k.a sparseness) is a hurdle in either case which we will refer to below. FLSs are high dimensional and hence their visualization is challenging, too. In this thesis we will mainly use WT-centric representation in which the WT is in the middle, surrounded by N_1, N_2, N_3, \dots as rings (See for example section 4.2 on page 26). This visualization lacks information about the connections between different rings, but we hope that along with other Fig.s it will be clearer.

1.2 Evolutionary Dynamics on Fitness Landscapes, The Quasi Species Theory

Introducing Fitness Landscapes without presenting prominent dynamical models dictating adaptive trajectories upon it is lacking. The simplest population dynamics upon FLSs can be expressed by the replicator equation:

$$\dot{x}_i = x_i(f_i - \bar{f}) \quad (2)$$

where x_i is the frequency [$\sum_{i=1}^n x_i = 1$] of genotype i in the population, f_i its fitness and $\bar{f}(t) = \sum_{i=1}^n x_i(t)f_i$ is the average population fitness. In this model the frequency of each genotype increases (decreases) if its fitness is higher (lower) than \bar{f} . The fittest takes over the population as time progresses (Fisher's Fundamental theorem of natural selection [8]). This model, incorporating only selection, i.e. without mutation, might be a good approximation for short periods of time. However, on the course of millions of years, mutations play a major evolutionary role (major drive for genetic variability) and can't be neglected. Incorporating mutations into the model leads us to the Quasi Species (QS) model (a.k.a replicator - mutator equation) which is

$$\dot{x}_i = x_i Q f_i + \sum_{j \neq i} x_j q_{ji} f_j - \bar{f} x_i \quad (3)$$

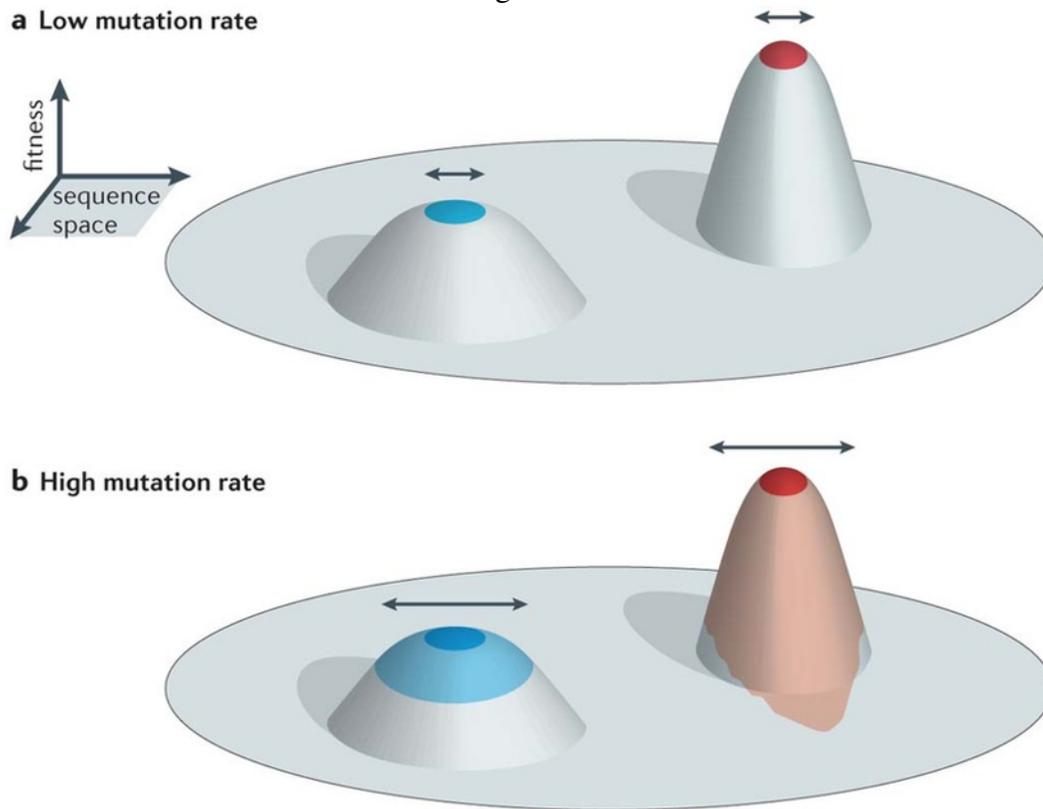
where Q is the quality coefficient - the probability to replicate with no error and

$$q_{ji} = (1 - q)^{L - h_{ji}} q^{h_{ji}}$$

q_{ji} is the transition probability (mutation rate) from genotype j to i where L is genome length, h_{ji} is the Hamming distance between genotypes j and i and q is the mutation rate per base pair. The further away genotype i is from j the less probable the transition is. If $q = 0$ and $Q = 1$ Equation 3 reduces to Equation 2. Note that while Equation 2 is deterministic, Equation 3 introduce stochasticity to the process, manifested by q and Q which are only the expectations of real life mutation rate distributions. This highlights some of the difficulties in predicting evolutionary dynamics.

While equation 2 results in Darwin's Survival of the fittest, in a population governed by the replicator-mutator dynamics, the population reaches mutation-selection balance, which could result in sub-optimal genotypes. Previous studies have shown [39] that robustness to mutations (a.k.a genotype flatness) can be evolutionary advantageous in a high mutation rate regime. Wilke et al. simulated evolutionary dynamics of a population of digital organisms upon a FLS with only 2 types of peaks: high and steep one and a low but shallow. Their results showed that as long as the mutation rate q (per base pair for that matter) was low enough, the population kept populating the high steep peak in great numbers (See Fig. 3 on the next page adopted from Lauring et al. [15]). However, as mutation rate crossed a certain critical value (denoted as q_c) the cost, due to mutations, that organisms on the high steep peak had to endure was so severe that it killed them, resulting in a population shift towards the low shallow peak. There the cost was durable, even though the fitness was lower. This phenomenon was coined "Survival of the Flattest" as opposed to Darwin's "Survival of the Fittest". Another outcome, due to the relative fitness indifference of the peak's occupying genotypes, was the co-existence of

Figure 3:



A. Lauring et al, 2013

Nature Reviews | Microbiology

Survival of the Flattest Effect and the Quasi-Species model. a) When the mutation rate is relatively low, the fittest genotype will prevail even if it is located on a steep peak. b) When a high mutation rate is in place, the population will spread out upon the FLS. The population located on the steep peak will pay a heavy price for each mutation, result in a shift - the shallow blue hill will be populated by various neighboring genotypes.

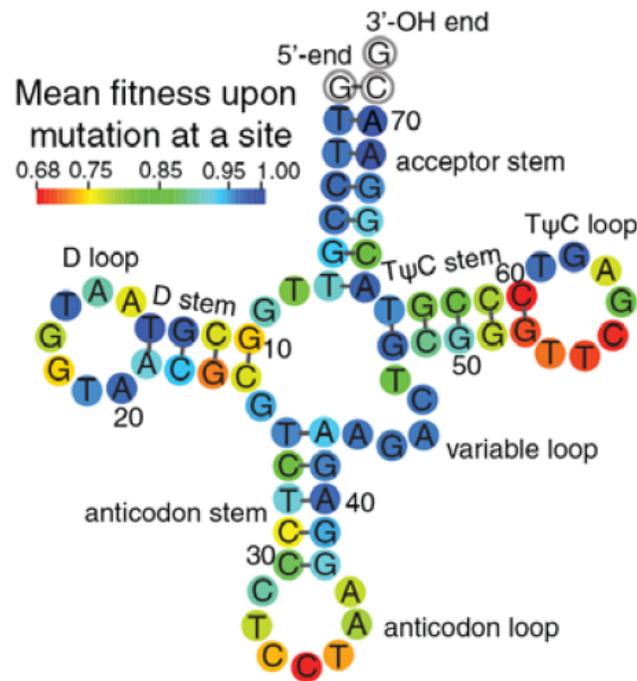
different close-by neighbors at equilibrium which later were referred as the “Quasi Species” (QS). Increasing the mutation rate even further crosses yet another threshold known as the error-threshold (denoted as q_e) above which the population collapses (For further reading see [23] chapter 3), i.e. all genotypes are equally populated irrespective of fitness.

To complete the picture, a study by Sardanye’s et al. [30] derived a simple formula for the critical mutation rate, above which Survival of the Flattest is in force, q_c , using Mean Field Theory (MFT) simplification modeling:

$$q_c = 1 - \left(\frac{f^{flat}}{f^{steep}} \right)^{1/L} \quad (4)$$

assuming, as in Wilke et al. study, only a two peak FLS. a steep one with fitness f^{steep} and a shallow one with fitness f^{flat} . Moreover, Nowak approximated the error threshold to be $q_e \approx \frac{1}{L}$ ([23] chapter 3).

Figure 4:



The secondary structure of $tRNA^{CCU-Arg}$. The yeast gene is non-essential and was Chuan et al. Fitness Landscape object of study. Here, we show the sensitivity to mutations of different regions in the tRNA such as the anticodon, B-Box (i.e. $T\psi C$ loop and stem) etc. Adopted from their paper[17]

1.3 Chuan et al. tRNA FLS Study

Recently, as mentioned above, a few studies that measured FLSs were published. The tRNA study of Chuan et al. was particularly attractive for analysis for several reasons:

- Their library consisted of an impressive amount of variants. 65K variants in the first study and 23K (a subset of the 65K) in the second.
- The tRNA's short length made it a better candidate than other studies [31, 27] (sparseness was less severe).
- In their later study, they measured fitness in 4 different environments.
- They measured Darwinian fitness (as opposed to others who used different fitness proxies)

Their pipeline for producing the data was as follows. They first created numerous mutated variants of the WT $tRNA^{CCU-Arg}$ strain using error-prone PCR where, of course, the WT performed as template. This yielded an impressive local coverage surrounding it: They reached almost all 1-neighbors ($N_1=99\%$. not 100% due to technical reasons), a large fraction of all the 2-neighbors ($N_2=60\%$) and also a few thousand variants from N_3 , N_4 and so on up to N_9 . They later transformed them into yeast cells with their WT tRNA gene deleted, to create the library. The next stage was to compete the pool of transformed variants against each other for 24 hours,

diluting 1/100 of the population after 12 hours. They performed the competition in 4 different conditions (in their later study): 23C, 30C (ideal yeast growth condition temperature), 37C and DMSO (oxidative stress) and sequenced the pool of variants in the beginning (T_0) and in the end of the competition (T_{24}). Only variants with more than 100 reads in both time points were considered for further analysis. Finally, they used fold-change analysis with respect to the WT strain, i.e. the fitness of a variant strain i_m in environment m was defined by -

$$Fitness(i_m) = \left(\frac{R_{24}^{i,m}/R_0^{i,m}}{R_{24}^{WT,m}/R_0^{WT}} \right)^{1/G^m}, \quad (5)$$

where $R_t^{i,m}$ is the read count of variant i , in environment m at time t , $G^m = \log_2 \left(d \cdot \frac{(\#WT \text{ cells at env } m \text{ at } T_{24})}{(\#WT \text{ cells at env } m \text{ at } T_0)} \right)$, $d = 0.01$ is the dilution coefficient. Hence, $Fitness(i_{WT}) = 1$ in each environment. This is correct when assuming only exponential growth phase along the 24 hours. In that case, Chuan et al. reported fitness $f_{i,m}$ at environment m equals $e^{r_{i,m}}$ where $r_{i,m}$ is the growth rate of variant i . It can be extracted using $r_{i,m} = \ln(Fitness(i_m)) + r_{WT,m}$ (see section 6.1 on page 45 for more details). Note that they set all fitness values lower than 0.5, to be 0.5 to avoid genetic drift effects. They repeated the competition in each of the conditions at least 3 times, to validate the measurement accuracy (See Fig.s - 45 through 48 on page 66 for correlations between repeats in each condition).

In the following chapter we will first state our work's goals and outline our hypothesis. Next, we will present major results in two consecutive chapters. the first one will discuss further insights into Chuan et al. data and will constitute as a platform for the following chapter that will try to address our main questions and goals. We will conclude with the discussion, comparing our results to other studies and stating open questions for further research.

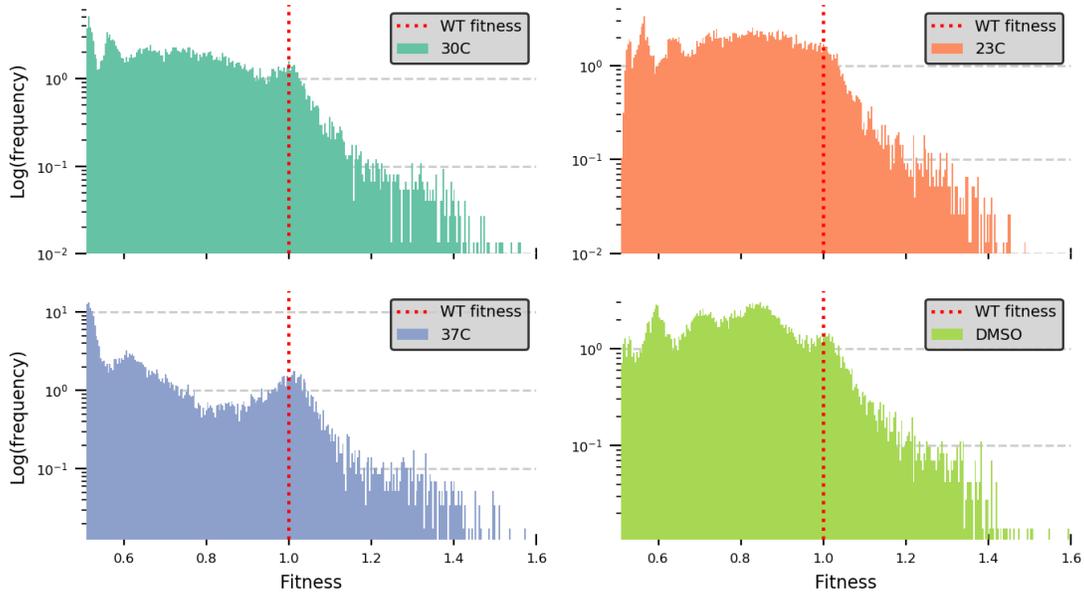
2 Goals

Below we show that the WT $tRNA^{CCU-Arg}$ gene isn't the fittest. In fact, more than 2.5% of the measured variants have higher fitness values. The main question that this work addresses is: Why is it so? What can explain the WT identity? What dynamics-governing factors could have made the WT who it is? Of course, along the work we raise further questions (e.g. see Discussion part on page 43) and answer others, but these questions were what we focused on.

We hypothesize that if the WT is not the fittest (in terms of growth rate) in any of the conditions, then it might be due to its flatness properties as defined in 1.2 on page 7 and in 4 on page 24. To do this, we first found a FLS in which the WT is sub-optimal, then we tried to identify distinctive WT features. Most notably are its fitness across many conditions, its fitness relative to its nearest neighbors, how much it is a local optimum and types of flatness. We finished with properly simulating evolutionary dynamics upon the given FLS, looking for evidence of a preferred (hopefully realistic) settings in which the WT prevails.

Asexual evolutionary dynamics is responsible for major biological processes, from bacteria and fungi to cancerous cells. Nevertheless it is still poorly understood, let alone predictable. This research tries to partially address and revisit basic scientific questions in this regard: Based on what properties genes get selected? What are the necessary conditions for that? How repeatable is this? We hope that the conclusions of this study can serve as a foundation for further research addressing these issues.

Figure 5:



Significant proportion (more than 2.5%) of the genotypes are fitter than the WT. Semi log scale distribution of fitness values in 4 different environments, based on Zhang’s data. In each of the 4 environments the WT (red dotted line) is not the fittest. The WT’s 1 and 2 neighbors are the major contributors to the high concentration surrounding fitness equal 1 (See Fig 15). This insight is invariant to reasonable read count deviations and not due to pre-competition background (trans) mutations (See subsection 3.2)

3 Results - Characterization of Chuan et al. Data

3.1 The WT is not the Fittest

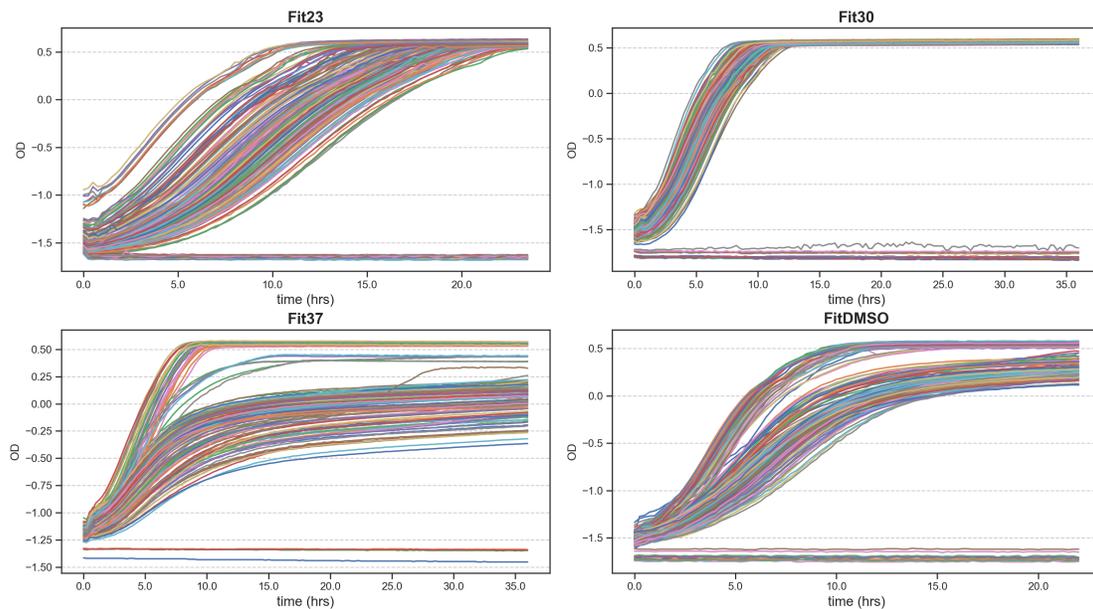
There is one thing striking about Chuan et al. data set, beside the reasons supplied above: The WT tRNA strain is a mid-fitness variant. Fig. 5 presents a semi-log scale distribution of the fitness values measured in each of the 4 environments. We find that a significant proportion of the genotypes (absolute amount of ~650, varied slightly between conditions, with fitness larger than 1.1 which comprise about 2.5% of the total variant pool) have higher fitness than the WT in each of them. This contradicts the intuitive guess that the WT is the fastest growing strain (fittest in this regard) which could suggest that it must have some other evolutionary advantage that led for its prevalence.

3.2 The Validity of the Experimental Observation

The WT being significantly less fit than relatively many other strains made us wonder why it is so and whether or not it is an experimental artifact. Chuan et al. did try to further confirm their en-mass competition results, without addressing this specific peculiarity. They estimated the fitness of 55 distinct randomly chosen genotypes in two more ways:

- Fitness estimation from growth curves of the 55 strains grown separately. Two biological replications of

Figure 6:



Variability in all 3 growth phases. OD growth curves of different 55 strains that were grown by Chuan et al. in all 4 environments. It is clear that some strains don't grow at all. Strains that do grow, vary in the exponential phase growth rate, in yield value and in the length of the lag phase. Growth rate variability also varies between conditions.

fitness measurement were performed per genotype.

- Fitness estimation by pairwise competition of the 55 strains. Three biological replications of fitness measurement were performed.

Moreover, they used a t-test to examine if the fitness of a variant is significantly different from 1 at a nominal P value of 5%. Although reassuring, this doesn't specifically confirm the above observation. First and foremost, it is crucial to understand the causes that might make one variant grow faster than another. Higher growth rate in the exponential phase is probably one of the main reasons. Nevertheless, if it were the only one, a variant with reported fitness of 1.3 would have an unrealistic doubling time of less than 60 minutes (See Materials & Methods). Short lag phase and late yield phase are other major causes as well - the variability in all 3 phases between OD growth curves of the 55 randomly sampled strains at the 4 different conditions shows that (See Fig. 6). Hence, having such higher fitness variants is not far fetched (even other studies reported similar results [16]) thus passing a first sanity check.

There are still two major concerns we found worth addressing in this regard:

1. The accumulation of beneficial background mutations (in the genome, not in the tRNA gene itself) during library preparation (which takes approximately 50 generations) that gave evolutionary advantage to the fitter-than-WT strains. In other words, fitter-than-WT genotypes grew faster not since their $tRNA^{CCU-Arg}$ sequence was better adapted but due to mutations elsewhere.

Figure 7:

#	tRNA	Length	Identity	View In GtRNAdb
1	Tetrapisipora_phaffii_CBS_4417_tRNA-Arg-CCT-1-1	72 bp	72/72 (100%)	View
2	Saccharomyces_cerevisiae_tRNA-Arg-CCT-1-1	72 bp	72/72 (100%)	View
3	Saccharomyces_sp_boulardii_ATCC_MYA-796_tRNA-Arg-CCT-1-1	72 bp	72/72 (100%)	View
4	Saccharomyces_sp_boulardii_17_tRNA-Arg-CCT-1-1	72 bp	72/72 (100%)	View
5	Lachancea_kluuyveri_NRRL_Y-12651_tRNA-Arg-CCT-1-1	72 bp	72/72 (100%)	View
6	Eremothecium_cymbalariae_DBVPG#7215_tRNA-Arg-CCT-1-1	72 bp	72/72 (100%)	View

$tRNA^{CCU-Arg}$ **Wild Type Gene is indeed wild.** Exact matches of the gene found in other Wild Species, based on GtRNAdb. It strengthens our premise that this tRNA gene was optimized by natural conditions and evolution and not by artificial ones present in the lab.

2. Read count sampling noise can also be a potential source of fitness estimation error, especially when the read count is low. Indeed, Chuan et al. only kept genotypes with only >100 reads at T_0 , but in the worst case when under-sampling at T_0 and over-sampling at T_{24} one might still end up with exceptionally high erroneous fitness measure.

Put differently, there might have been an error in fitness estimation or alternatively the fitness was correctly measured, but it wasn't due to tRNA mutation. Before treating these two concerns hereunder we only note that yet another concern might be that the Wild Type is not wild to begin-with, i.e. that that strain mainly evolved during lab evolution and that is the reason there are higher fitness variants than it. We dismiss this concern by simply showing that other wild species share the same gene sequence (based on GtRNAdb, see Fig. 7).

3.2.1 Background Mutations

To mitigate the concern that fitter-than-WT genotypes grew faster not since their $tRNA^{CCU-Arg}$ sequence was better adapted but due to mutations elsewhere, we will roughly estimate the chance (order of magnitude) for such an event to happen. We will only focus on fitter double mutants, since by Fig. 8 below they comprise the majority of the variants with fitness higher than the WT. Given such a high fitness variant, denote its fitness to be \tilde{f} . The concern put differently, it might be that this value is a weighted average of a mixture of two types of organisms:

1. Variant i that differs from the WT only in two tRNA nts (a **real** double mutant).
2. The same variant as the first, with a beneficial BG mutation (a **disguised** double mutant). Denote it as j .

This means that variant i might not be fitter than the WT after all. We first claim that unless variant j constitutes most of the mixture, there isn't much room for concern. To show this, we break it down into 3 possibilities, that differ by the mixture composition:

Table 1: Different Mixture Composition Outcomes

	High % of real tRNA variants	~ equal %	High % of disguised variants
Disguised variants are fitter	Disguised variants are unrealistically fit.	Either disguised variants are unrealistically fit or that the real tRNA variants' fitness is ≥ 1	See further analysis of this scenario below.

The cases where the real tRNA is fitter or that both have similar fitness lead to the same conclusion: The real tRNA variant fitness is larger than \tilde{f} . The first two scenarios from the left are unrealistic. For example, in the first let the percentage of the real tRNAs be 80%, and assume they have similar fitness as the WT. For the disguised variant to “create” an alleged 1.3 fitness organism it should be 2.5 times fitter (!) than the WT. If the real tRNA fitness is less than the WT than it is even more so, and if its larger than the WT then our observation left intact. Thus we will only estimate the probability of the third one.

In Chuan et al. experimental procedure, they first synthesized the different variants using error prone PCR initiated with the WT strain and then transformed its output variants into yeast cells (See [17, 18] SI for more details). We estimate that only a small fraction (say 0.001%) of the cells was transformed with a variant. Since there are only $O(10^3)$ different types of double mutants variants in the experiment, it is reasonable to assume that only one copy from each of these double mutants transformed into cells. This is also the worst case we can consider, shown in table 1.

Next we need to estimate the probability for background mutations during the transformation stage which usually lasts 50 generations. For that we will use the results of a study made by Levy et al. [16] in which they quantitatively tracked the evolutionary dynamics of half a million lineages of yeast cells, each tagged by a unique bar code. One of their remarkable findings was the spectrum of beneficial mutation rates as a function of the fitness effect which is the added percentage to the growth rate ($\ln(\text{fitness}_i)$ in Chuan et al data, and see Materials & Methods). They report that the beneficial mutation rate per cell per generation that leads to at least 5% fitness increment is $O(10^{-6})$. Hence we assess the expectation for our $O(10^3)$ double mutants yeast cells to have a beneficial mutation in the first generation to be on the order of 10^{-3} . For a background mutation to yield the same fitness measurement, only now for a mixture (50%-50%) of real and disguised mutants, it has to be at least 2 times more beneficial. Since the spectrum of beneficial mutation rates reported by Levy et al. is decreasing exponentially in part, flat in part but drops sharply near 12% fitness effect, we estimate such a beneficial mutation to be far less likely, and thus negligible. This leaves the expectation of background beneficial mutations during transformation at 10^{-3} , far less from the 2.5% beneficial mutations found in Chuan et al. data. While at most ~10% of the fitter genotypes could be explained by background mutations, this can not fully explain the measurements.

We draw another reinforcement to our conclusion from BLASTing our tRNA against other fungi species. If some of the fitter tRNA sequences are evolutionary advantageous, we expected to find them in other types of

fungi. Most of the BLAST results were missing in the Chuan et al data, i.e. variants didn't exist, but some did. The most interesting one was a double mutant with a 1.51 fitness score at 30c found at *Eremothecium sincaudum*. This suggests that at least some of the variants with higher fitness than the WT are real and not disguised.

Finally, we performed yet another control in which we searched for shared properties among genotypes with fitness higher than 1 that might explain why it is so. If their fitness effect is due to beneficial tRNA mutations, it might have a signature on the secondary structure of the tRNA molecule. Fig. 8 presents the normalized (after dividing by the size of each neighborhood) distribution of the variants with fitness higher than 1.1 which is clearly dominated by N_2 genotypes, while for variants with fitness lower than 0.9 there is no such trend (see Fig. 9). Although this might suggest that their high fitness is due to better Watson-Crick complementarity, we excluded that as a reason since only less than 2% of them were so. We next tried to localize the mutations upon the tRNA secondary structure, normalized by that region size and to itself (by dividing by the overall amount of variants to reach values between 0 and 1. see Fig. 10). Indeed, most of the mutations didn't occur in sensitive places as the B-Box or the anti-codon (ac) (See also Fig. 4 on page 9 for tRNA secondary structure map) in high fitness variants. When comparing to mutation locations of lower than 0.9, it seems that beneficial mutations have a slight tendency to cluster in the acceptor stem and main loop (χ^2 test, p-value < 0.002 for all conditions). With that, in Fig. 11 we looked at nt usage to see if there is a certain preference for high fitness mutations. They do seem to be Thymine (T) rich, especially when compared to mutations with fitness lower than 0.9 (χ^2 test, p-value < 10^{-8} for all conditions). It is interesting to note here that the T nt (which turns to U) has the largest amount of known Post Transcription Modifications (PTMs) and thus might make the tRNA more "modifiable" and flexible. Overall we do see some distinctive properties shared among the high fitness variants which reinforce our conclusion, namely that background mutation are not the reason for higher than the WT fitness variants, even further.

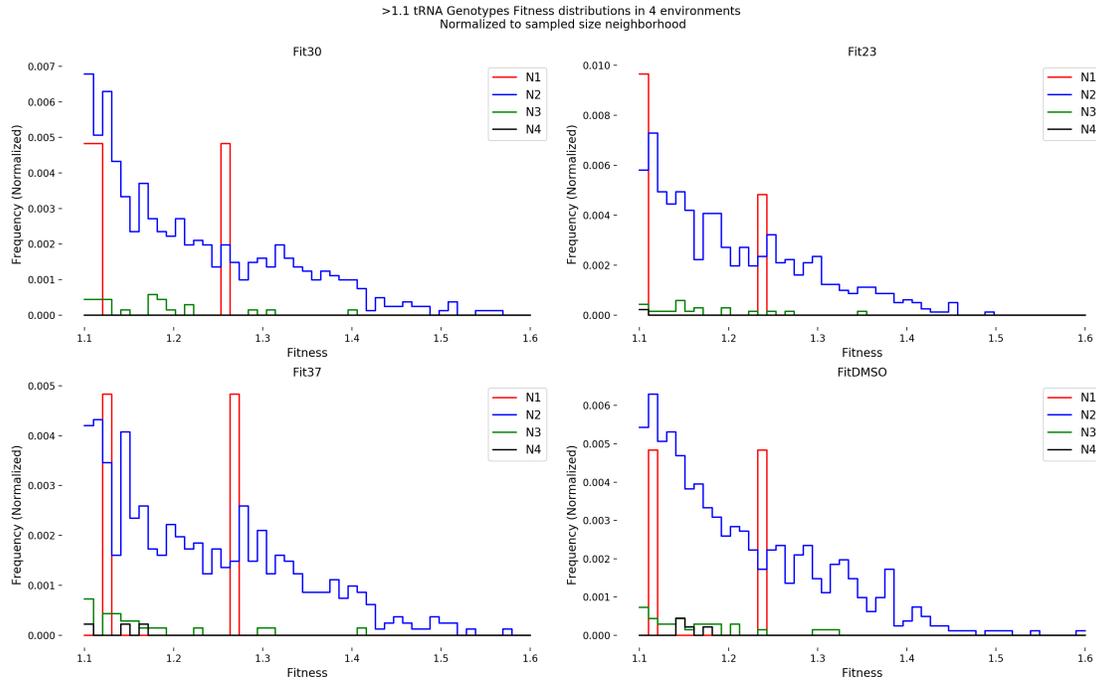
3.2.2 Read-Count Sampling Noise Effect

In this part we address our second concern, namely that that the high fitness values Chuan et al. measured were due to sampling noise. Under-sampling at T_0 (the time when the competition between the variants started) and over-sampling at T_{24} (the time it ended) one might still end up with exceptionally high erroneous fitness measure.

Fig. 12A is a 2D histogram presenting the $\ln(\#T_0\text{reads})$ for each of the fitness values in the experiment. As mentioned above, the relative low number of read counts at T_0 of variants with fitness higher than 1 raises a concern and might cause fitness misassessment. We used bootstrapping to estimate the fluctuation in the number of reads sampled. More specifically, we assumed that the sampled read count distribution at T_0 is the true experimental distribution of cells (which, of course, is false), and simulated multiple sampling of 10% of it, each time. The result in standard deviation of sampling used us to calculate the possible fitness range, which then manifested in the confidence intervals shown in Fig. 12B.

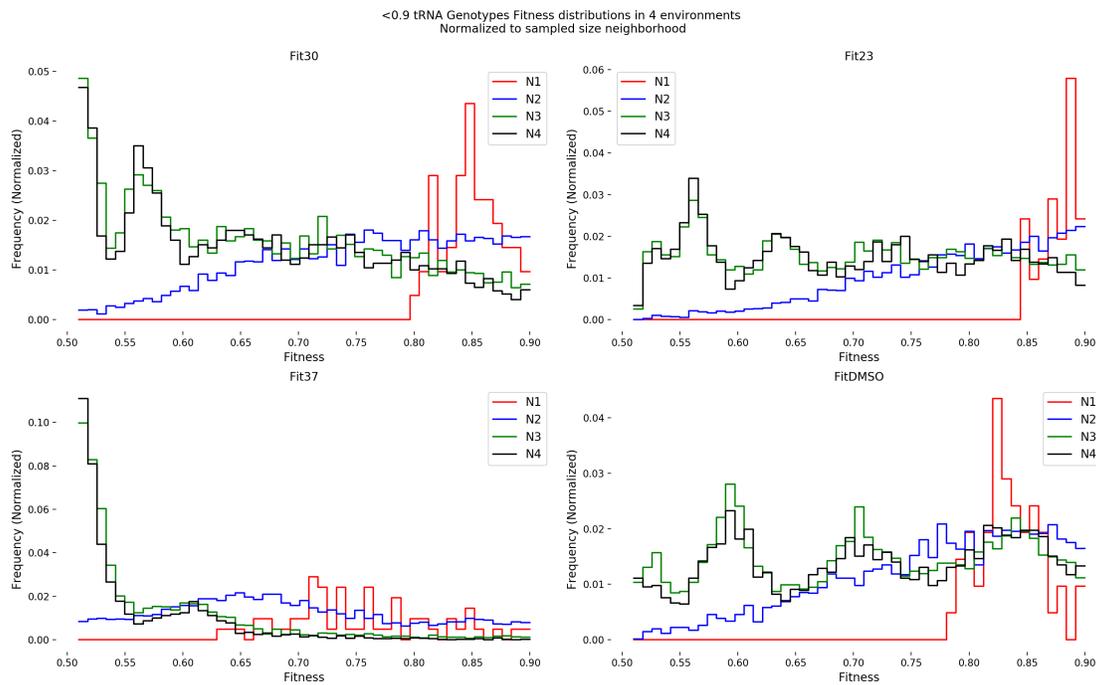
In another statistical check, we used the standard deviation of the 6 replications as a measure of the sampling noise. The standard deviation there yielded yet another confidence interval as shown in Figs 13 and 14. Even

Figure 8:



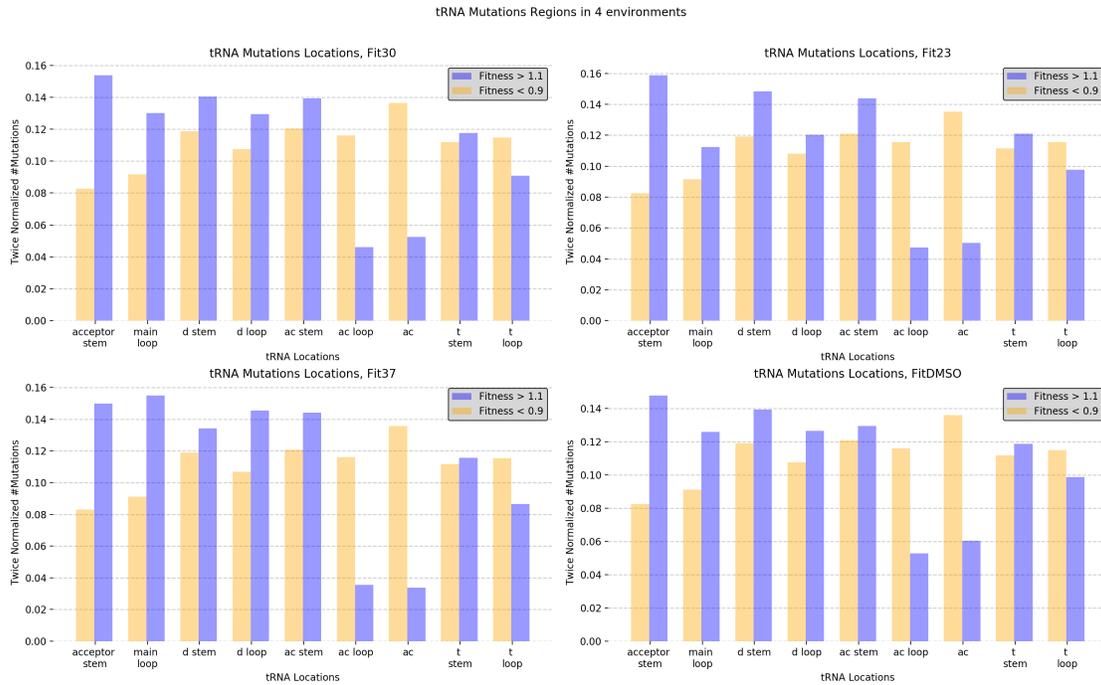
The majority of the fitter-than-WT genotypes are double mutants. Fitness histogram of the absolute number of variants, normalized by the size of the relevant neighborhood, in all 4 conditions. N_2 genotypes are the most abundant among variants with fitness larger than 1.1, even after taking the size of the neighborhood into account.

Figure 9:



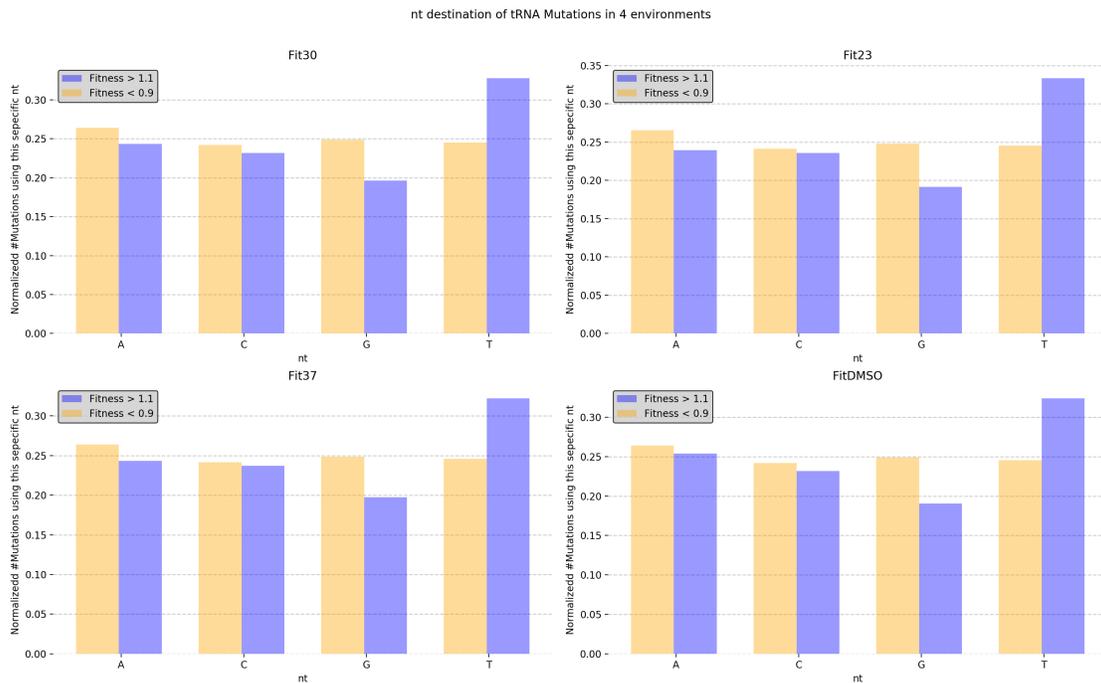
Low fitness variants are populated mostly by N_3 and N_4 . Fitness histogram of the absolute number of variants, normalized by the size of the relevant neighborhood, in all 4 conditions.

Figure 10:



Higher and lower fitness mutations locations in tRNA structure in all 4 environments. The y axis is normalized to itself and to the tRNA region length. There is an under representation of mutations in anti-codon (ac) and ac loop among high fitness variants, suggesting a functional role. When comparing to mutation locations of lower than 0.9, beneficial mutations seems to have a slight tendency to cluster in the acceptor stem and main loop (χ^2 test, p-value < 0.002 for all conditions).

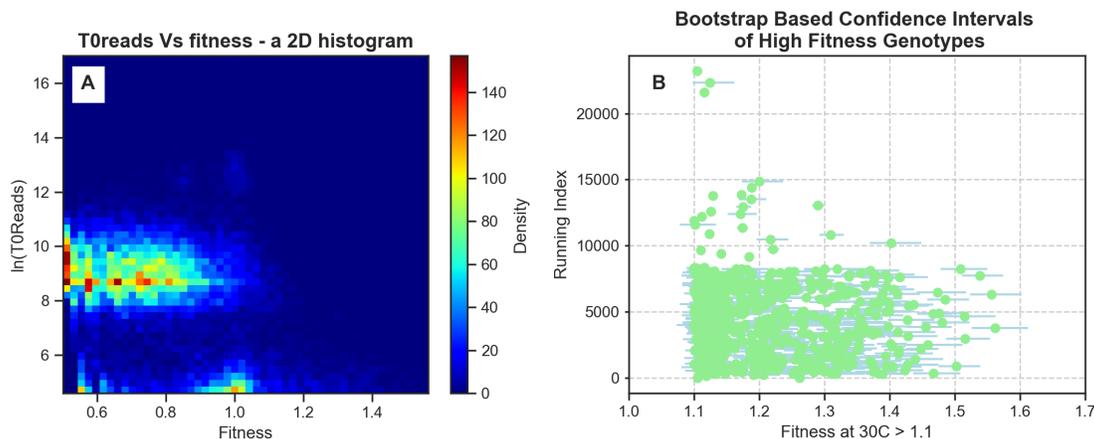
Figure 11:



Fraction of nucleotide usage in mutations of high and low fitness variants in 4 environments. When comparing to low fitness variants, higher ones tend to mutate more to the Thymine (T) nt, together with an under representation of G (χ^2 test, p-value < 10^{-8} for all conditions).

Figure 12:

Confidence of high fitness variants



Confidence of high fitness variants using bootstrapping. A) 2D histogram of fitness values (at 30C) against log of the number of reads of each variant at the beginning of the competition (T_0). Relatively low read count of variants with fitness higher than 1 raises a concern. Namely that they were subject to sub-sampling noise and hence their fitness was miscalculated. B) Confidence bars calculation using bootstrapping. Put together with Zhang’s careful examination of the competition results and Fig. 13, we dismiss this concern.

though the standard deviation between the biological repetitions is higher, in both cases, more than 90% of variants with fitness higher than 1 still stay that way, leaving our observation valid.

3.3 The ‘Anna Karenina Effect’ and Data Sparseness

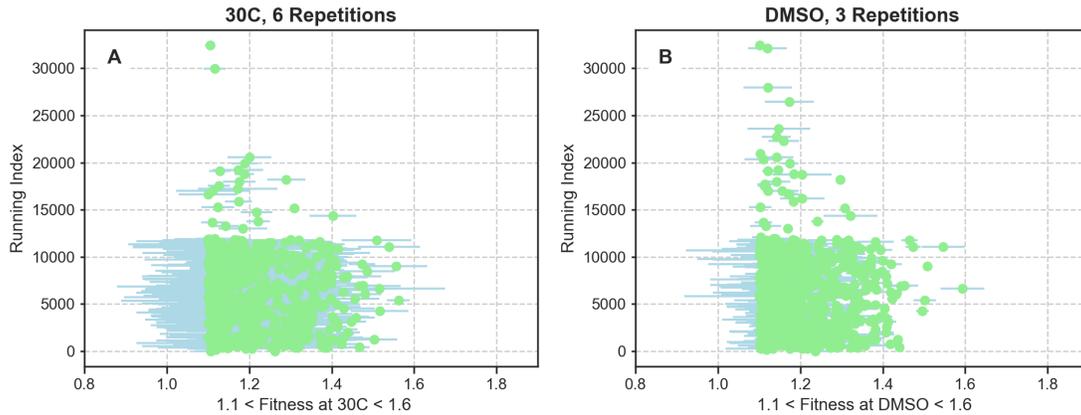
Further exploration of Chuan et al. data reveals interesting things. Fig. 15 is a breakdown of Fig. 5 histogram to the different WT’s neighborhoods in the 4 different conditions. The high amount of variants surrounding fitness equal 1 is not due to the WT’s N_1 variants but thanks to its 2-neighbors. Moreover, not surprisingly, farther genotypes from the WT tend to be less fit with the exception of the double mutants. 37C condition seems to give rise to a bigger variance of fitness values already at N_1 signifying how hard it is for the tRNA variants to remain functional in high temperatures.

One could argue that while the WT is never the fittest, there is no single genotype which fitter in all 4 conditions. To rule out his possibility we performed correlations between different intervals of fitness values (See fig 16 and 17) and concluded that the superior in fitness genotypes tend to keep their relative supremacy when conditions change, while the inferior genotypes don’t. We called that the “Anna Karenina Effect” - “Happy families (superior genotypes) are all alike. every unhappy family (inferior genotype) is unhappy in its own way” (Anna Karenina opening sentence). This phenomenon suggests that environment change can be a big evolutionary drive. It can dramatically change a poor fitness genotype, which later, with mutations, can reach a more stable superior genotype that won’t degrade even after environmental change.

We end this chapter by noting what a mere fraction Chuan et al.’s data comprise, out of the entire tRNA FLS. The amount of all nt sequences of length 72 is $4^{72} = 2^{144} \approx (2^{10})^{14} \approx (10^3)^{14} = 10^{42}$, where the sampled data

Figure 13:

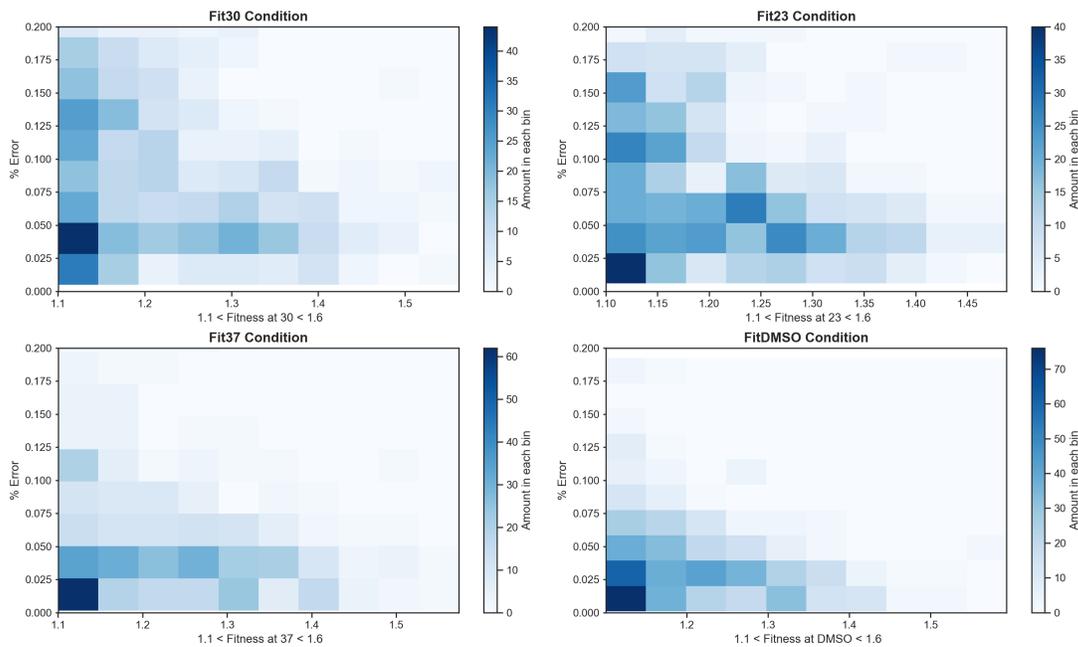
Confidence of high fitness variants,
Based on biological repetitions; 30C & DMSO



Confidence intervals of high fitness variants based on biological replications. A) Reported fitness values at 30C against some running index, with a symmetric confidence interval added to each based on the standard deviation of 6 different biological replications. B) The same, only for DMSO condition, based on 3 replications. Both panels emphasize the conclusion made in Fig. 12 - most (more than 90%) of variants fitness stay well above WT's fitness of 1. Fig 14 present these results in a different way.

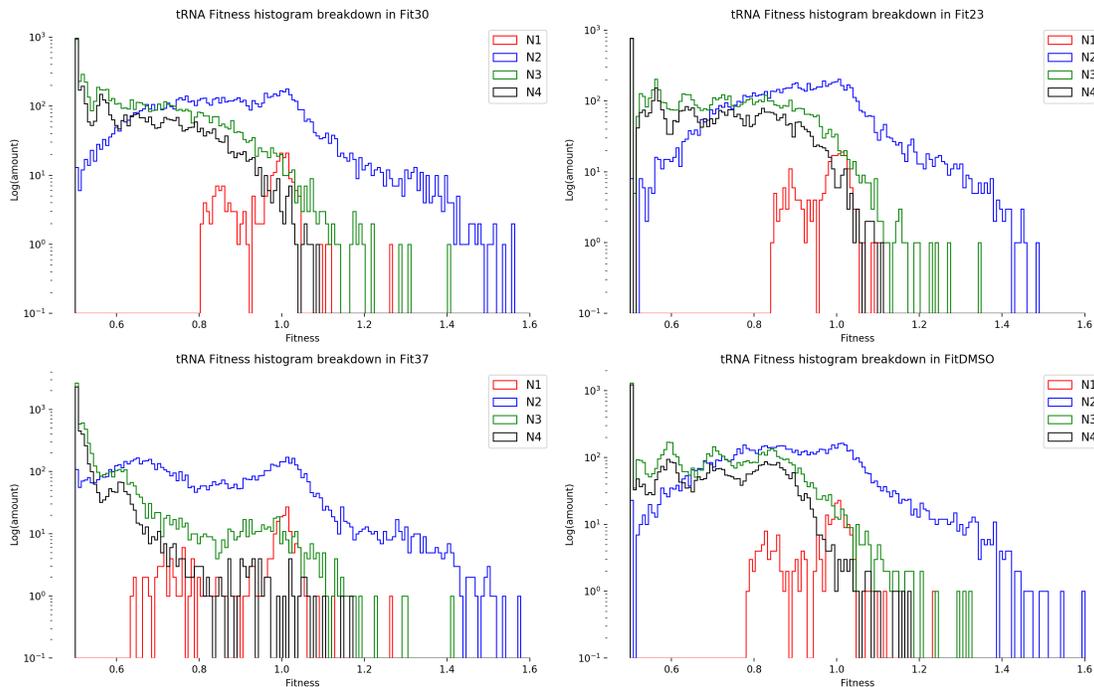
Figure 14:

2D histograms of %Error vs Fitness in the 4 conditions, based on the 6 repetitions



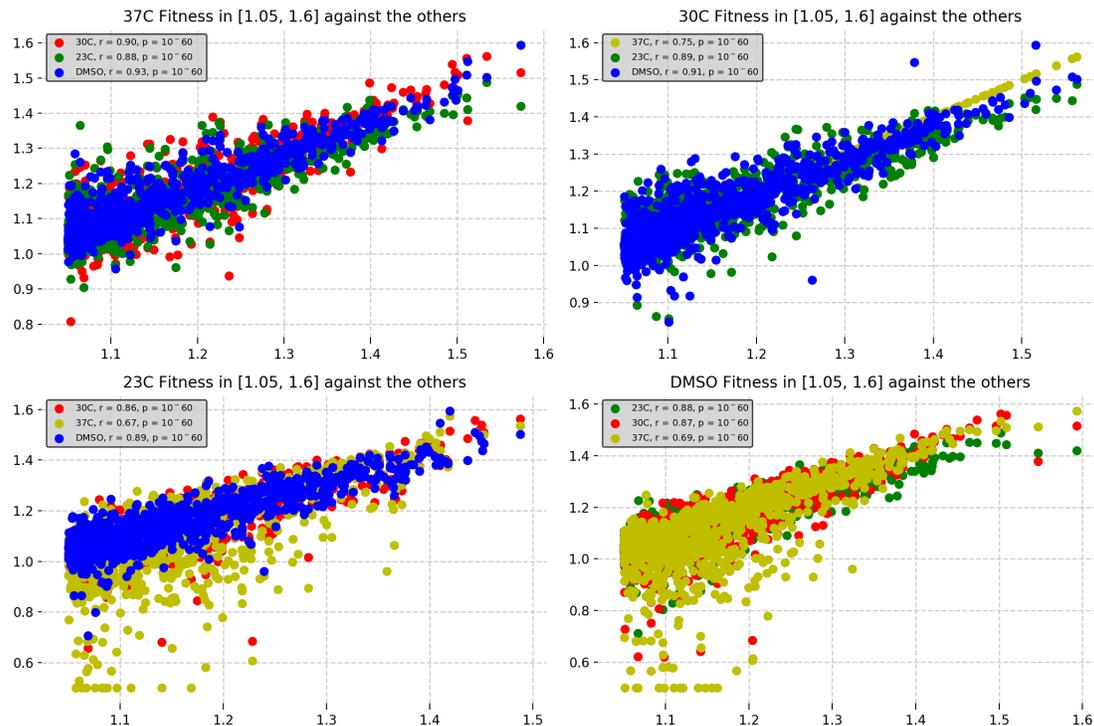
2D histogram of %Error VS Fitness of high fitness variants, 4 conditions. There are only a handful of variants (less than 10%) with %error (which is the standard deviation of measured fitness values in the different biological replicates) above 10%. In other words, less than 10% of the high fitness variants can be explained by noisy read count sampling.

Figure 15:



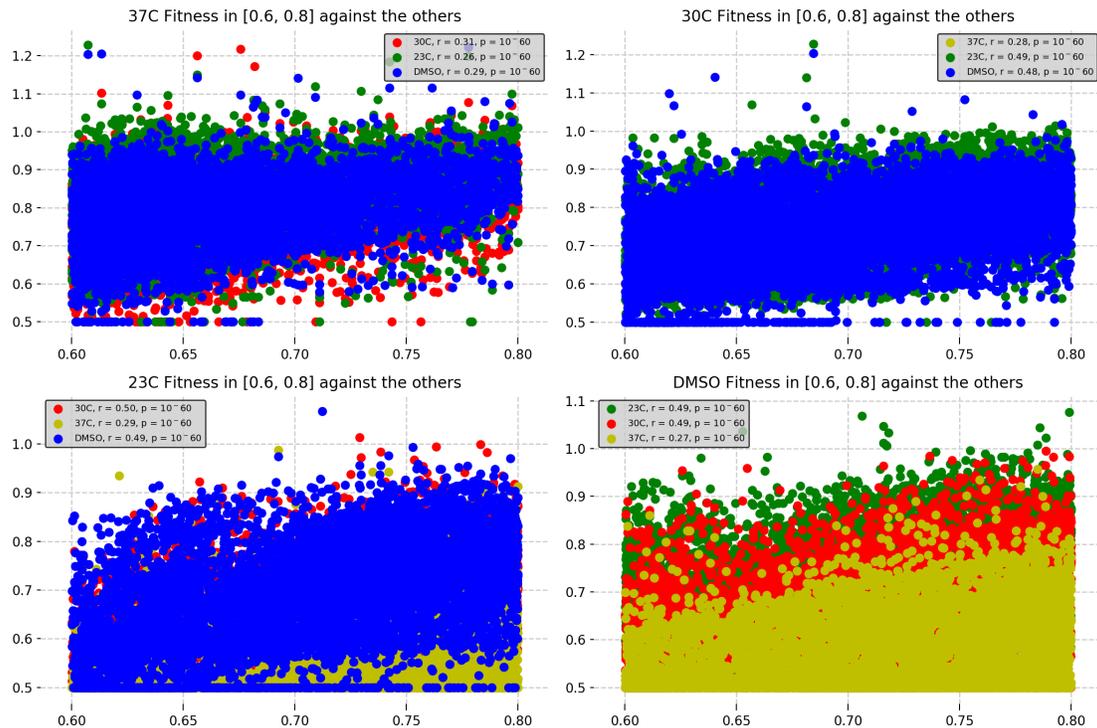
Fitness values histogram breakdown according to different neighborhoods. Semi log scale histogram of fitness values in 4 different environments, breakdown according to Hamming distances from the WT. N_1 and N_2 are the major contributors to the high concentration surrounding fitness equals 1.

Figure 16:



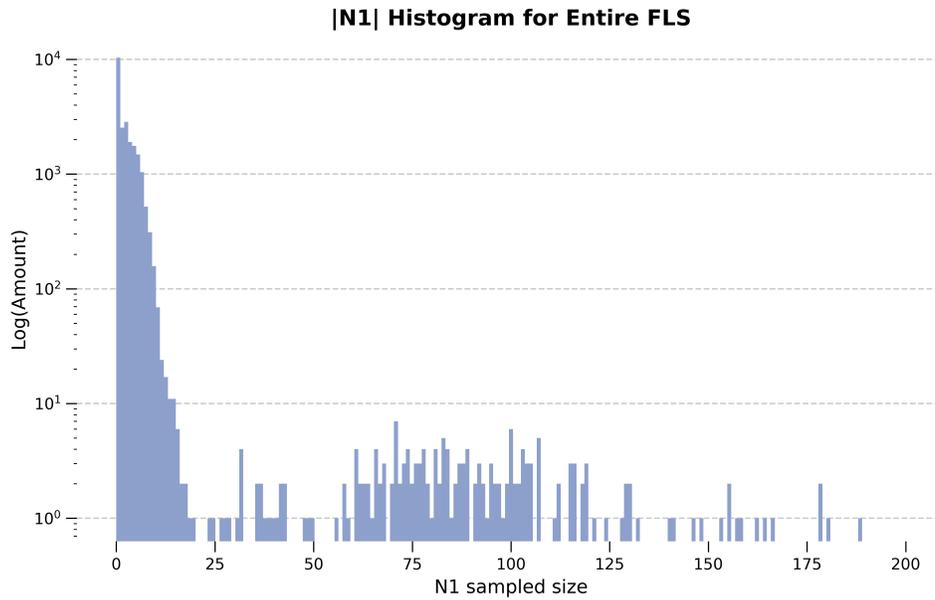
High fitness variants stay high after condition change. The 'Anna Karenina Effect'. We correlated high fitness values in each condition against their values in all other conditions. The relatively high correlation coefficient (see legend) indicates that the high fitness variants stay high after conditions change.

Figure 17:



Low correlation between the fitness values of low fitness genotypes, across conditions. The 'Anna Karenina Effect'. We correlated low fitness values in each condition against their values in all other conditions. The poor correlation coefficient (see legend) indicates that each low fitness variant has its own unique fate after conditions change.

Figure 18:



Extreme partially sampled sequence domain. Sampled 1-neighbors histogram for all existing variants. One can witness the sub-sampling influence. Generally, the number of k -neighbors of a sequence of length 72 equals $3^k \binom{72}{k}$, hence for 1-neighbors it is $3 \cdot 72 = 207$. We tried to tackle this problem statistically in various ways.

is of the order 10^5 . This is of course a lot in absolute terms, especially when comparing to similar data sets, but only a very small portion ($\frac{10^5}{10^{42}} = 10^{-37}$) of all possible genotypes. Therefore it is inevitable to sub sample the FLS, result in a sampling bias towards the WT. Fig. 18 presents a semi-log scale histogram of the size of N_1 for all the variants (not just the WT) sampled. Most of the variants have no neighbors present in the data, while the others have a small amount. We addressed this issue in various ways which we will describe later, but note it is a built-in property of experimental FLSs with no escaping.

4 Results - Looking for Wild Type's Evolutionary Advantages

There are several plausible answers that can explain what makes a mid-fitness genotype a WT. In this chapter we will address several of them using FLS data, starting by rejecting the fraudulent and less probable followed by our attempt to keep and reinforce others.

4.1 Wild Type is (not) the fittest across conditions

One possible explanation that comes straight to mind is that even though the WT is not the fittest in each condition alone, perhaps it is fittest on average. In our setup, fitness values measurements in each condition yield a different FLS mapping. Previous studies suggested (See for example [8] chapter 3) that when the evolutionary time scale is much slower than the times it takes for the environment to change, as in our case, one might treat the FLS upon which selection is acting as some combination (Gillespie in the above book modeled it as a geometric mean, $f_e = (\prod_{m=1}^M f^m)^{1/M}$ where f_e is the effective fitness, m signify different conditions) of the FLSs per condition. In other words, selection acts upon an effective FLS, computed from the original FLSs in each condition.

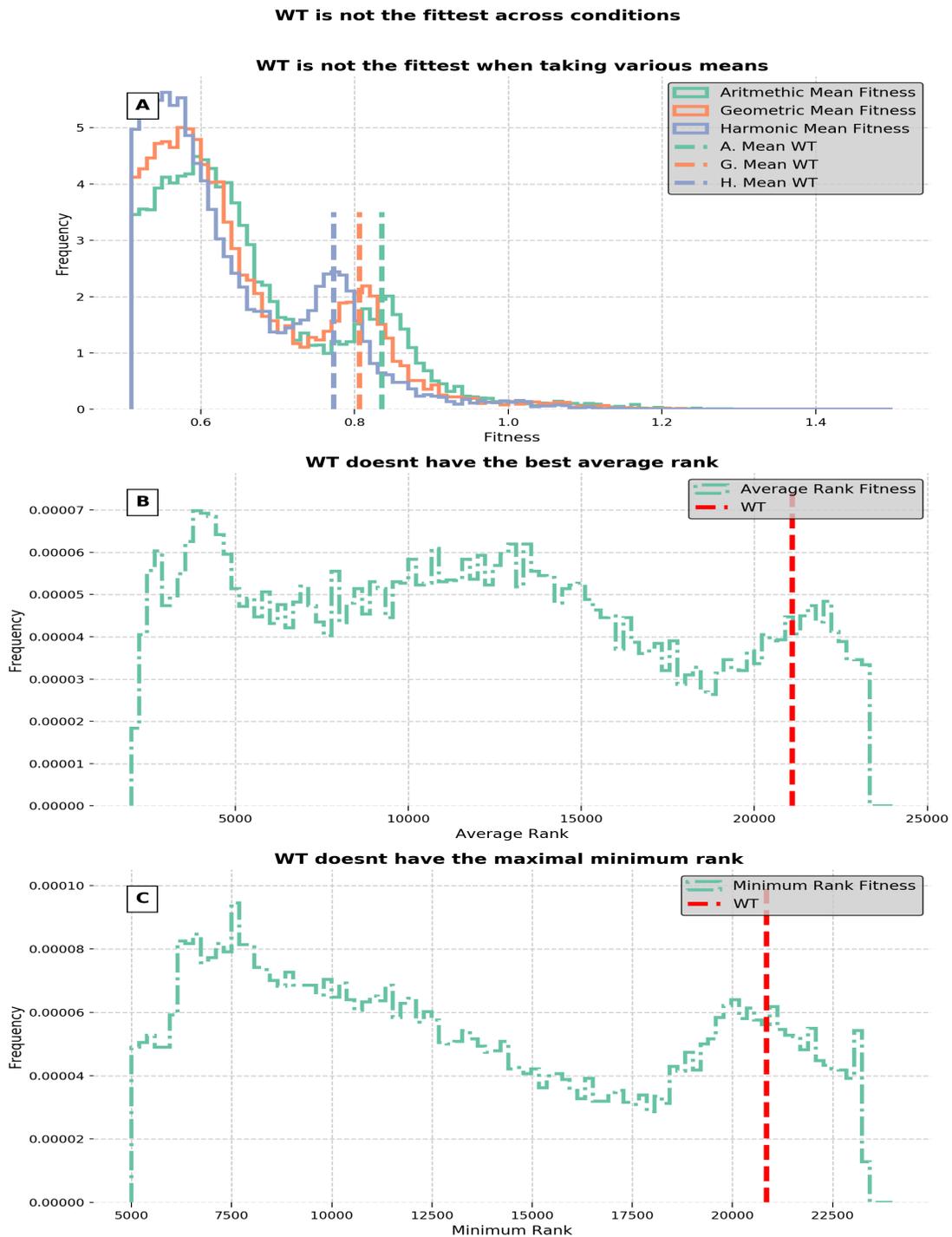
Computation of effective FLS can be done most naively using means of different kind. However, remember that our fitness values are relative to the WT in each condition, whose fitness was set to be 1. Thus, before we calculate the mean, we must align fitness values of different conditions for them to be comparable. To do that, we used Chuan et al. reported WT growth rates for each of the conditions (For the exact calculation see Materials & Methods, subsection 6.1 on page 45). In Fig. 19A we plotted the histogram of the effective FLS values after averaging the aligned original values using arithmetic, geometric and harmonic mean (the latter can be thought of as a lower bound). The WT is not the fittest in any of the means calculated, i.e. not the fittest across conditions.

This result is not surprising given the 'Anna Karenina' effect described above. For the WT to be fittest on average, genotypes with fitness higher than 1 in some conditions must be lower than 1 in others, and that is not the case. The only caveat in this calculation is the underlying assumption that the weight corresponding to each environment is equal. Yeast cell most probably spends more time at 30C than at both extremes, 23C and 37C. Since we don't know the best way for weighing each environment, we may look at the WT's rank.

In Fig. 19B we see that the WT's rank (the position in which the WT's fitness value is in after increasingly sorting all the fitness values) is significantly inferior in all 4 conditions. Maximizing across genotypes the minimal rank across conditions might also be advantageous, for it can be considered as a sort of bottleneck, incurring the most damage when relevant. Nevertheless, Fig. 19C shows that it's not the case either - the WT didn't manage to maintain highest minimal rank (Maximal Minimum).

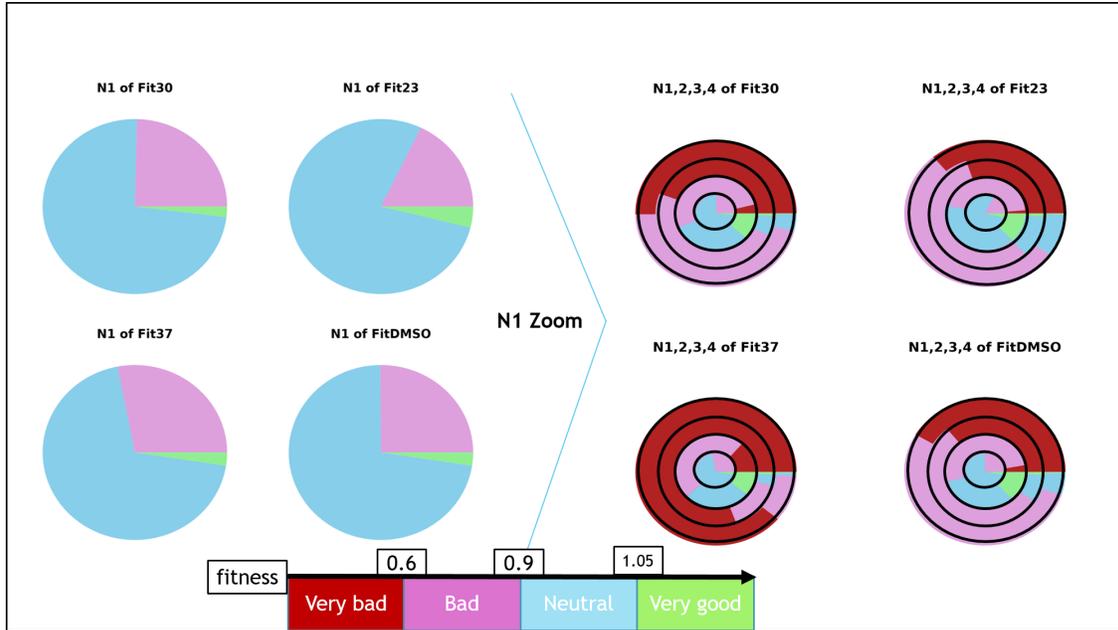
Of course, there might be other more significant conditions that weren't measured in which the WT is indeed fittest which can hamper the conclusions of the entire analysis. However, measuring fitness at additional conditions is very expensive, and hence for the lack of data we reject this explanation in the meantime.

Figure 19:



WT is not the fittest across conditions. A) Distribution of tRNA variants' fitness values averaged across the 4 environments in 3 different ways. B) Distribution of tRNA variants' rank (in terms of fitness values) averaged arithmetically across the 4 environments. C) Distribution of tRNA variants' minimal rank (in terms of fitness values) taken over the 4 environments. In each of the panels it is clear that a significant proportion of the genotypes are fitter (only in this regard) than the WT.

Figure 20:



The WT is a 'semi' local maximum. WT-centric visualization of the FLS in 4 conditions, surrounded by 4 neighborhoods (N_1, N_2, N_3, N_4). Colors signify fitness values. Once more - the majority of fitter-than-WT variants are double mutants. Zooming in to each of the 4 environments' N_1 , we find a subset, albeit small, of genotypes which are fitter-than-WT. Hence it is safe to say that the WT is a 'semi' local optimum.

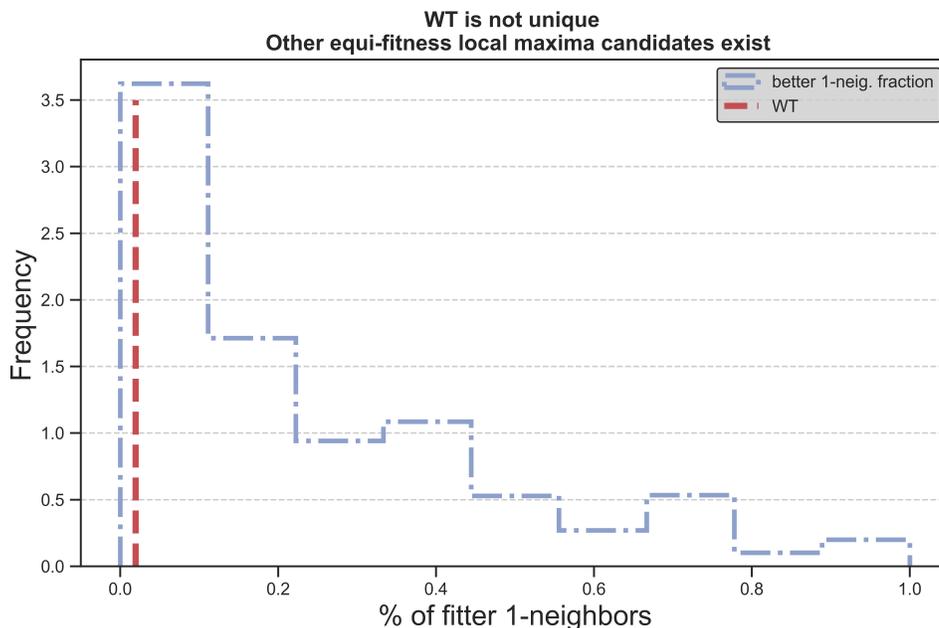
4.2 Evolution (didn't) 'get stuck'

Another possibility is that there are indeed genotypes fitter than the WT, only that the evolutionary process didn't get a chance to visit them yet. That might be the case if it got stuck on a local maximum with very low probability of getting out of. To explore this option we assessed how much of a local maximum is the WT. Fig. 20 shows a WT-centric FLS visualization in each of the 4 conditions: The WT at the center, surrounded by the the first few neighborhoods (N_1, N_2, N_3, N_4), shaped as rings. Fitness values ranges are color-coded as shown. Zooming to the first neighborhood (N_1) is convincing - the WT is a 'semi' local maximum with only a mere fraction of its 1-neighbors (i.e. genotypes in N_1) superior in fitness. However, analyzing equi-fitness variants reveals it is not something unique to the WT.

Fig. 21 shows that a large proportion of variants with fitness values between 0.85 to 1.15 have a similar percentage of superior 1-neighbors as the WT. In other words, there are many other possible 'semi' local maxima which are just as good. Hence, there is nothing special about the WT in this regard. Still, the stochastic nature of evolutionary dynamics might end up in such a way that made the WT yeast strain land on that specific peak of all others. It doesn't rule out the possibility that evolution got stuck.

To fully convince ourselves that indeed there is a relatively high probability for evolution to leave the WT region we used the available fitness landscape data once more to map trajectories of strictly increasing fitness, originating from the WT. Since most of the superior variants are double mutants, (See Fig. 20 or Fig. 8 on page 17) we calculated all the 2-step increasing-in-fitness trajectories. We found about 1000 of them, made possible due to the small amount yet meaningful fitter 1-neighbors (in other words if the WT was a complete

Figure 21:



The WT is not unique. There are many other equi-fitness semi-local maxima. We calculated for each variant with a fitness value between 0.85 to 1.15 at 30C, the percentage of its fitter 1-neighbors (i.e. the size of the green circle sector shown in Fig. 20). Here we present a histogram of that percentage and see that there are other semi-local maxima variants. Hence, the WT is not unique in this regard.

local maximum - such trajectories wouldn't exist). This amount comprises 2% of all possible 2-step trajectories originating from the WT which equals $(69 \cdot 3)^2 \approx 4 \cdot 10^4$. When we consider a realistic 10^8 yeast population size, over the course of time, it is unlikely for the evolutionary process not to leave the WT (assuming equal probability for all trajectories). Reassuring is the fact that this calculation is only a lower bound if taking neutral (non-decreasing) trajectories into account. To conclude, evolution most probably 'visited' these superior mutants, which in turn didn't manage to survive as opposed to the WT.

Note that the sub-sampling bias presented in Fig. 18 on page 23 is only beneficent with the above conclusions. It is much harder to be a local maximum with many more 1-neighbors surrounding each variant. Further, the sub-sampling is another reason for the amount of ascending trajectories calculated above being just a lower bound, i.e. they probably comprise more than 2% which makes it even more probable for evolution to realize these superior mutants.

4.3 The WT is robust to mutations (Genotype Flatness) in each condition

The question remains unanswered: What can explain the WT survival, given that evolution most probably realized superior genotypes? Our search for an answer from here to the end of this current chapter will focus on the role of "flatness" in the above evolutionary settings. In our context, flatness is a synonym for robustness to changes of different origin. In this part we will examine the WT's relative endurance to mutations (sequence change). Motivated by "Survival of the Flattest" effect and QS theory, we ask whether the WT is exceptionally

flat, for it might be the reason for it's identity.

We will say that a genotype i is genotype-flat in the j 'th direction if $|f_i - f_j|$ is relatively small. This definition is equivalent to calculating the norm of the respective (discrete) directional derivative when treating FLS as a discrete realization of some high dimensional Euclidean space. Taking it a bit further, we will say that a genotype i is genotype-flat if, on average, it is relatively flat on all directions. Formally, we define steepness of genotype i as the average gradient between variant i and its nearest neighbors for which we have data:

$$s_i = \frac{1}{|N_1(i)|} \sum_{j \in N_1(i)} |f_i - f_j| = \frac{1}{m} \sum_{j=1}^m |f_i - f_j| \quad (6)$$

when m is the number of all of i 's 1- neighbors. All genotypes have the same number of 1-neighbors, but in practice the available data on each genotype's N_1 neighborhood is only partial (See Fig. 18 on page 23 for demonstration). Naming this genotype property flatness is a bit misleading - The higher the score s_i is, the more steep the genotype gets. Hence, we will refer to this property as steepness when it is understood that low steepness, i.e. flatness, is the feature of interest.

In order to assert whether or not the WT is relatively robust to mutations, we calculated the steepness score s_i for each one of the variants in Chuan et al. data set, wherever possible. In Fig. 22 we show steepness score distribution, breakdown by different fitness ranges at 4 conditions. The WT's steepness score at 30C for example is well below genotypes of fitness larger than 1.2 at 30C, on the lower 5% among the genotypes with fitness ranging between 1.05 to 1.2 at 30C and on the lower 10% for among genotypes with lower fitness values. We obtained similar results in the 3 other conditions as well.

Moreover, the correlation between steepness and fitness reveals another unique aspect of the WT's genotype-flatness (See Fig. 23). It maximizes the fitness to flatness ratio at 30C. In other words, the WT is the highest peak in the FLS that still has a relative flat surroundings. Again, this was repeated in all other conditions.

These results are also resilient (fig 24) when considering worst case errors based on the 6 replications (See Fig 13 on page 20). By worst case we mean reducing to the center variant fitness f_i one standard deviation due to sampling noise while to all of its surrounding 1-neighbors we add one standard deviation. Therefore we change equation 6 to be:

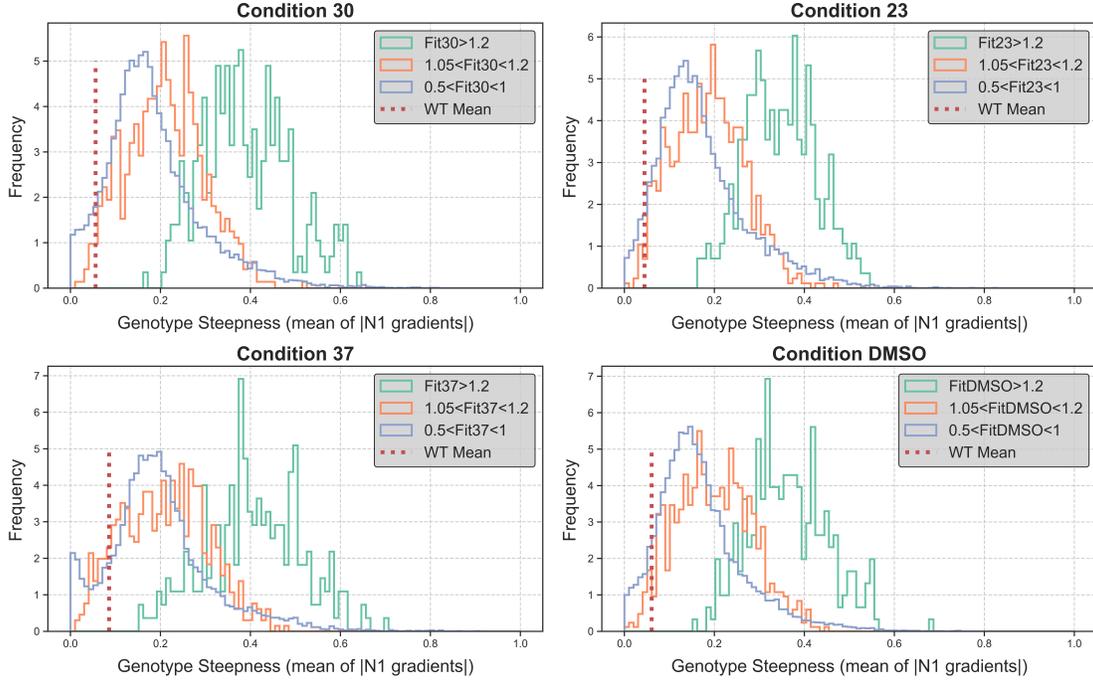
$$\hat{s}_i = \frac{1}{|N_1(i)|} \sum_{j \in N_1(i)} |(f_i - \sigma_i) - (f_j + \sigma_i)| \quad (7)$$

where \hat{s}_i is the new steepness, after taking into account sampling noise, and σ_i is the fitness standard deviation between the 6 biological replications.

4.3.1 Dealing with sampling bias

As mentioned above, it wasn't always possible or meaningful to calculate the steepness score of a variant due to the scarcity of available 1-neighbors. As some genotypes have very low known 1-neighbors, we considered only genotypes with at least 5 to 10 available 1-neighbors. We chose that threshold so on the one hand it will

Figure 22:

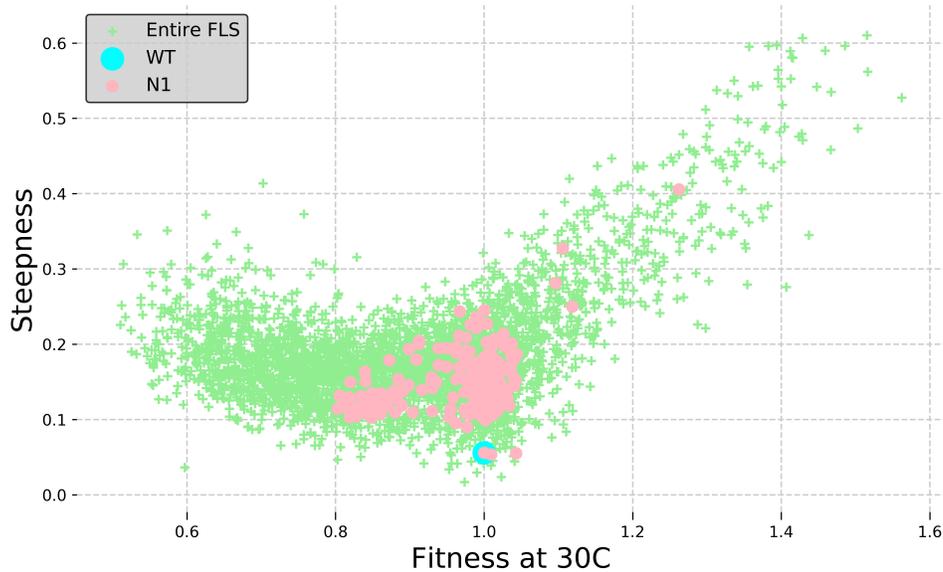


WT genotype flatness. Steepness distribution, breakdown by different fitness ranges at 4 conditions is shown. In all 4, the WT steepness is low relative to variants with higher fitness values. The WT's steepness score at 30C for example is well below genotypes of fitness larger than 1.2 at 30C, on the lower 5% among the genotypes with fitness ranging between 1.05 to 1.2 at 30C and on the lower 10% for genotypes with fitness values ≤ 0.5 .

be sufficiently high but on the other will not exclude most of the data. Additionally, another concern emerged. Namely that the WT flatness reported above is only an artifact of the way the data was 'sampled' (using error-prone PCR with the WT as original template): dense around the WT while sparse towards the outskirts. We controlled for that in the following way: we exhaustively went through all $|N_1|$ values of all the mutants in the data set with fitness larger than 1, and for each value, which we'll denote by m , we draw m 1-neighbors from the WT's N_1^{WT} repeatedly. We then calculated the steepness score (denote as \tilde{s}_{WT}^m) for each of the repetitions and compared their mean (denote as $\overline{\tilde{s}_{WT}^m}$) to the minimal steepness score of mutants with m 1-neighbors, which we will denote as $\min_j(s_j^m)$ (the minimum was taken across all variants with m neighbors with fitness larger than 1). The results are shown in Fig. 25 together with further details on the comparison. The WT's flatness is still convincing the moment we cross a very low amount of neighbors.

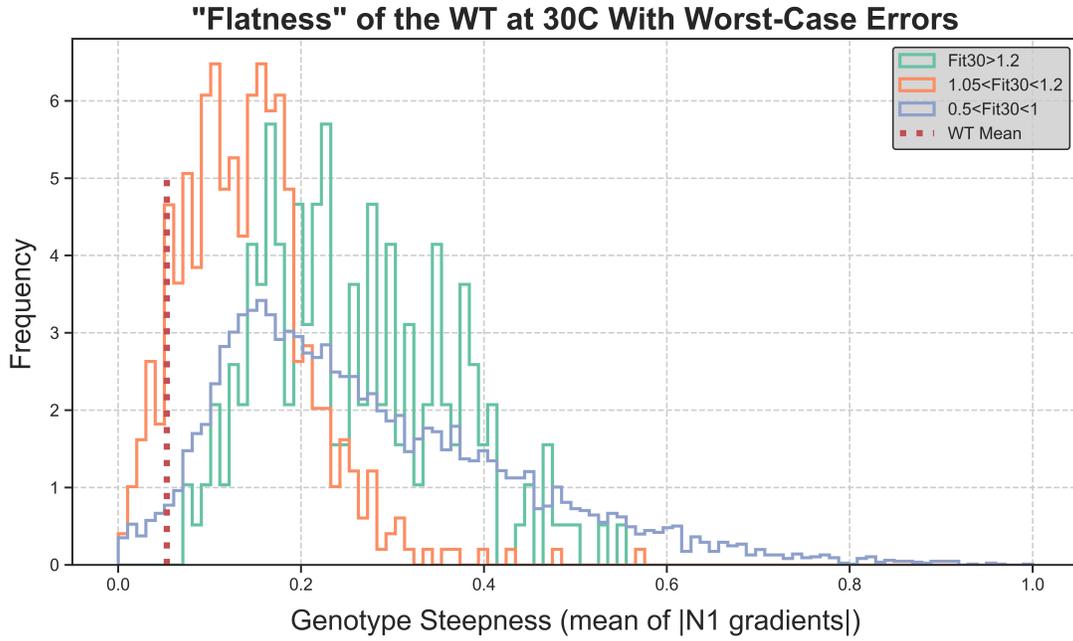
Figure 23:

Steepness VS Fitness Correlation Threshold = 5



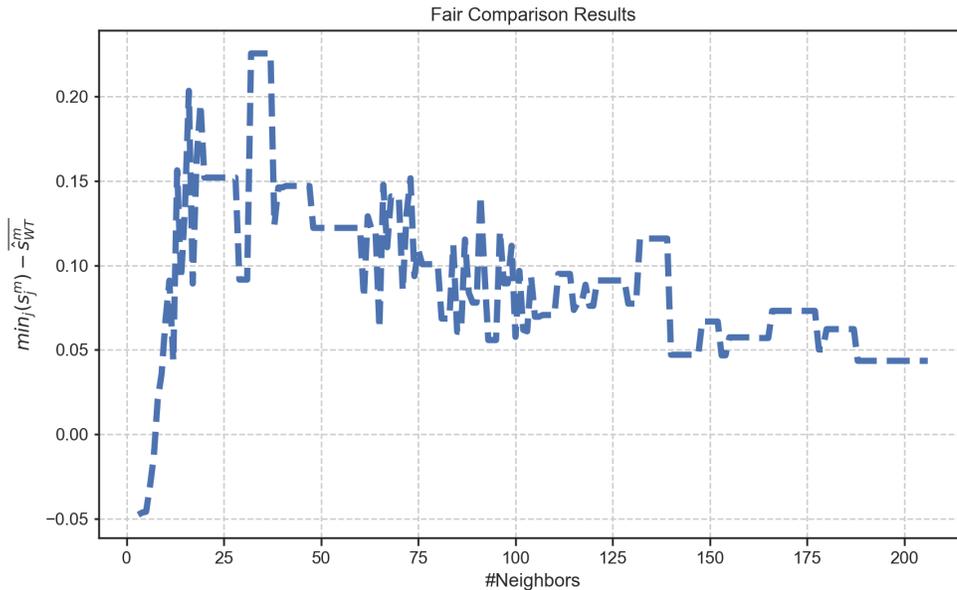
The WT maximizes the ratio of genotype flatness to fitness. Correlation of steepness against fitness, the WT and its N_1 neighborhood are highlighted in different colors. Only variants with more than 5 1-neighbors are presented. The WT and some of its 1-neighbors have relatively low steepness score compared to their fitness. The correlation coefficient between steepness and fitness with values between 0.9 to 1.6 is $r = 0.86$ ($p < 1\%$) and for fitness values between 0.5 to 0.9 is $r = -0.376$ ($p < 1\%$). $r < 0$ for the latter since there is a high probability for low fitness values to be surrounded by genotypes with higher fitness (i.e. to be situated in a valley). This is interpreted as high steepness as well due to our definition which used absolute value. This V shaped correlation disappears and becomes a straight line for steepness score calculated without the absolute value.

Figure 24:



Flatness of the WT at 30C under Errors. Similar histogram as in fig 22, different by the use of shifted steepness scores based on our fitness confidence intervals. By shifting the flatness score as in equation 7 we only make it more difficult for the WT to stay relatively flat, and indeed the steepness score distributions got closer to the WT when compared to Fig. 22. Still the WT remained relatively flat.

Figure 25:

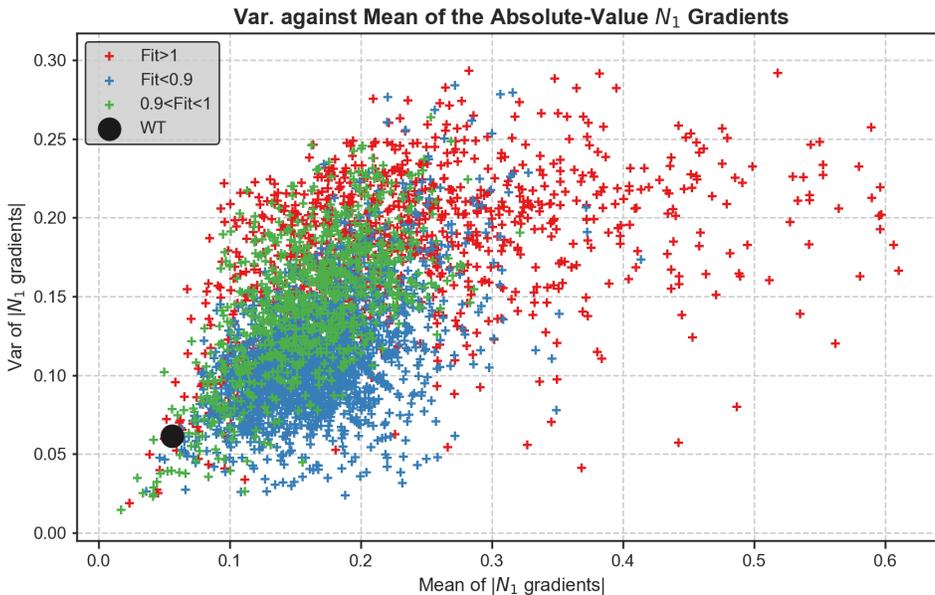


The WT is flat, even when considering sampling bias towards it. To control the allegedly WT's flatness (demonstrated in Fig. 22) for a possible artifact made by the sampling bias towards it, we devised a fair comparison process: we divided all variants with fitness higher than 1 to sets based on the number of their 1-neighbors $|N_1| = m$ exist in the data set. Then, for each set, we repeatedly sampled corresponding m 1-neighbors from N_1^{WT} and calculated their steepness \bar{s}_{WT}^m each time. Next we compared the mean over all repetitions, $\overline{\bar{s}_{WT}^m}$, to the minimal steepness score of each set, $\min_j(s_j^m)$, by subtracting. The figure presents the results of $a := \min_j(s_j^m) - \overline{\bar{s}_{WT}^m}$ against $|N_1| = m$, the number of 1-neighbors.

When a is negative, for example when $m = 3$, it means that there is a variant with 3 1-neighbors with a lower

In another related calculation, we plotted each variant's steepness s against the variance of its directional derivatives, after using Bessel Correction to account for the unequal sample sizes. The WT's relative low variance shown in Fig. 26 indicates high concentration around the mean, i.e. around the reported steepness score. Hence, with high probability, $\overline{s_{WT}^m}$ for a given subset of m directional derivatives will be very close to s_{WT} , strengthening the previous control. To conclude, our analysis shows that indeed the WT is flat.

Figure 26:



The WT has relatively low variance around its steepness score s_{WT} , compared to other genotypes. Each dot represents a different variant, showing its directional derivatives variance against their mean (its steepness score). The WT has one of the lowest variance and mean values relative to the rest of the data set. In other words $\overline{s_{WT}^m}$ is very well concentrated around s_{WT} .

The variance values presented account for unequal sample size. We corrected that using Bessel Correction: We multiplied each variance value by $\frac{n}{n-1}$ when n is the number of 1-neighbors.

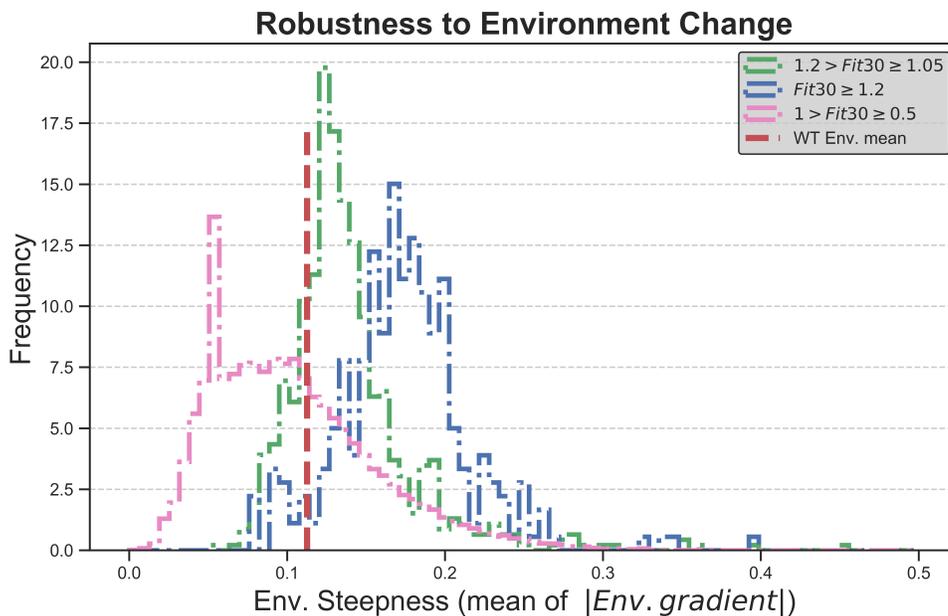
4.4 WT is robust to environmental changes (Environment Flatness)

Since the WT is robust to genotype changes (i.e. point mutations), we decided to check robustness to another change made possible by Chuan et al. data. In this part we show that the WT is also environmentally flat, i.e. its fitness does not change much upon changes of conditions. We performed the exact same analysis as in the previous section, with two major differences:

1. The directional derivative is taken now between fitness values of each particular genotype between different environments. Of course this derivative is meaningful only after aligning the values between the four conditions (See Material & Methods). We next average the 3 directional derivatives, choosing the fitness at 30C to be the center from which all of them are calculated:

$$s_i^{env} = \frac{1}{3} [|f_i^{30c} - f_i^{23C}| + |f_i^{30c} - f_i^{37C}| + |f_i^{30c} - f_i^{DMSO}|]$$

Figure 27:



Environmental steepness score distribution, breakdown by different fitness ranges is shown. The WT environmental steepness score is low relative to variants with higher fitness values. The WT's environmental steepness score is on the lower 3% among the genotypes with fitness larger than 1.2 at 30C and on the lower 15% among genotypes with fitness ranging between 1.05 to 1.2.

The underlying assumption of this definition is that all environments are equivalent (this is not a weighted average) and located equally distant from one another. Of course this assumption isn't true, but it is necessary for simplification. We couldn't think of a proper justification for using specific weights.

2. In this comparison there is no sampling bias and hence no concern of biased results - all the genotypes' fitness values were measured in all 4 conditions.

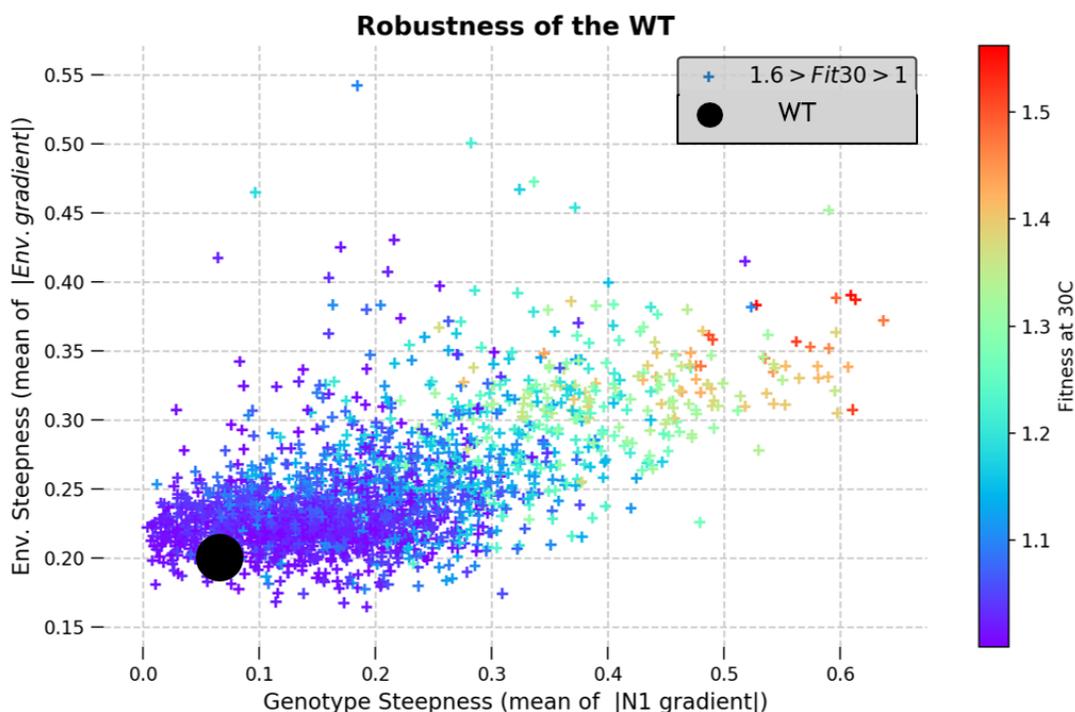
Of course, conclusions based on averages over 3 values are not that convincing, but given that FLS data measured in several conditions is hard to come by - we performed the analysis nonetheless. In Fig. 27 we show the same type of distribution as in Fig. 22, only this time portraying environmental steepness. The WT's environmental steepness score is on the lower 3% among the genotypes with fitness larger than 1.2 at 30C and on the lower 15% among genotypes with fitness ranging between 1.05 to 1.2.

Fig. 28 presents unified results for both flatness measures. We find that the WT is amongst the lowest in both types of flatness scores and thus relatively robust to both examined changes.

4.5 Does robustness grant evolutionary advantage?

So far we characterized the FLS structure. To relate structure properties to the evolutionary advantage of particular genotypes, we first need to study evolutionary dynamics in the given FLS. We were particularly intrigued by the WT sub-optimality. Hence we ran evolutionary dynamics simulations upon the given tRNA

Figure 28:



Robustness of the WT. Environmental flatness against genotype flatness for all genotypes in the data with fitness higher than 1 at 30C. The WT is amongst the flattest in both measures.

FLS to find whether or not there exist some settings in which the WT will prevail. For that, we used the Quasi Species framework and looked for the “Survival of the Flattest effect”.

We ran the simulation on a truncated FLS of ~15K genotypes (out of ~23K) consisting of at most 3 mutations away (including) from the WT, thus making sure there are no “islands” (genotypes that have no 1-neighbors in the data set). No “islands” policy is required to fulfill Peron-Frobenius Theorem assumptions, which in turn ensures existence and uniqueness of solution to the governing dynamic equations as well as convergence. We initialized the simulation randomly, using each time 10% of the overall possible genotypes, up to $N = 10^4$ organisms which was set to be the population fixed size. That way, each of the randomly chosen genotypes had roughly 6 (10K/1.5K) instances at generation zero. We ran the simulation at various environment conditions and mutation rates with at least 3 repetitions for each parameter set. See Materials & Methods for a more detailed description.

4.5.1 Simulation Results

In order to look for the Survival of the Flattest Effect, we kept track of three important values during the simulation run:

- Average population fitness, calculated using $\bar{f} = \sum_{i=1}^{15000} x_i f_i$, where x_i is the fraction of the i 'th variant and f_i its fitness.

- Population spread (variance) upon the FLS, calculated using entropy measure $E = -\sum_{i=1}^{15000} x_i \log(x_i)$.
- Population fraction which is in WT's N_1 (denote as N_1^{WT}) $R = \sum_{(i \in N_1^{WT} \cup WT)} x_i$ (i.e. sum of x_i over all variants in $N_1^{WT} \cup WT$)

If the Survival of the Fittest is in force, the population will have high average fitness, and thus will be densely centered on top of one of the high steep peaks. If and when Survival of the Flattest is in force we expect to see a shift to a lower shallow peak, hopefully N_1^{WT} peak in our case which is equivalent to an increase in R , a sharp decrease in average population fitness to around 1 (which is the WT's fitness) and an increase in population entropy E .

The results shown in Fig. 29 present indeed that. The three panels within portray \bar{f}, E, R as a function of the mutation rate q per nt per generation. Each point is the average over 3 repetitions of the respective value after a few hundred generations. When $q < 10^{-3}$ the survival of the fittest is dominant in all conditions, the population ends up densely located on a high peak while N_1^{WT} peak is relatively isolated. However when $q > 10^{-3}$ there is a transition. the average fitness drops to around 1 while E and R sharply increase. The population shifts towards the WT's shallow peak (See also Fig. 30- population composition in single runs). Therefore we have strong evidence to believe it's Survival of the Flattest effect that we see for $q_c \gtrsim 10^{-3}$. This value roughly agrees with Sardanye's et al. formula [30]: $q_c = 1 - \left(\frac{f^{flat}}{f^{peak}}\right)^{1/L} = 1 - \left(\frac{1}{1.5}\right)^{1/72} \approx 5 \cdot 10^{-3}$.

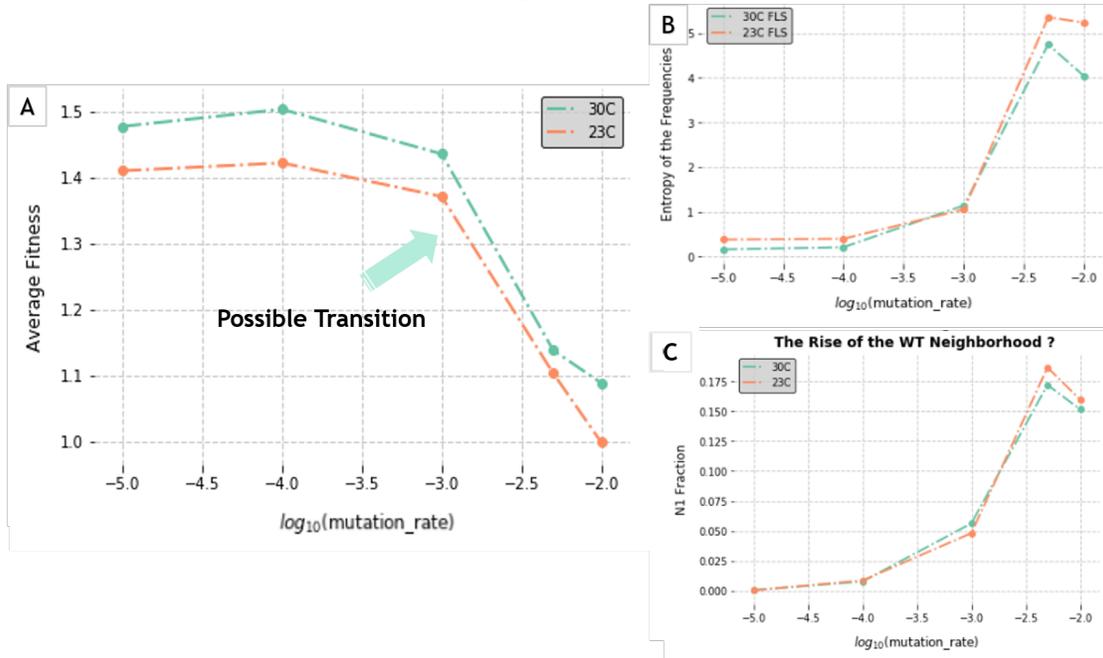
The reason we don't see a sharp transition in average fitness as in Adami et al. or Sardanye's et al. simulations [39, 30] is that in the latter they build a theoretical model comprised solely of flat and steep peaks. The only transition possible is between these peaks which are responsible for the abruptness. The reality is more complicated than that. Our simulation implements many different genotypes and uses an experimentally measured FLS with a more complex structure which give rise to a more complex, 'smooth' transition. As mutation rate increases, the probability to reach different genotypes and peaks increases as well.

4.5.2 Bridging Mutation Gap

Even though this result looks promising, two things overshadow the simulation's results. The first and maybe most apparent one is that the mutation rate q and q_c (which can explain survival of the WT) are as far as we know unrealistic. Previous studies [19] that mapped mutation rate per base pair in yeast DNA replication reported significant lower values of at most $10^{-8} - 10^{-9}$ per nt per generation.

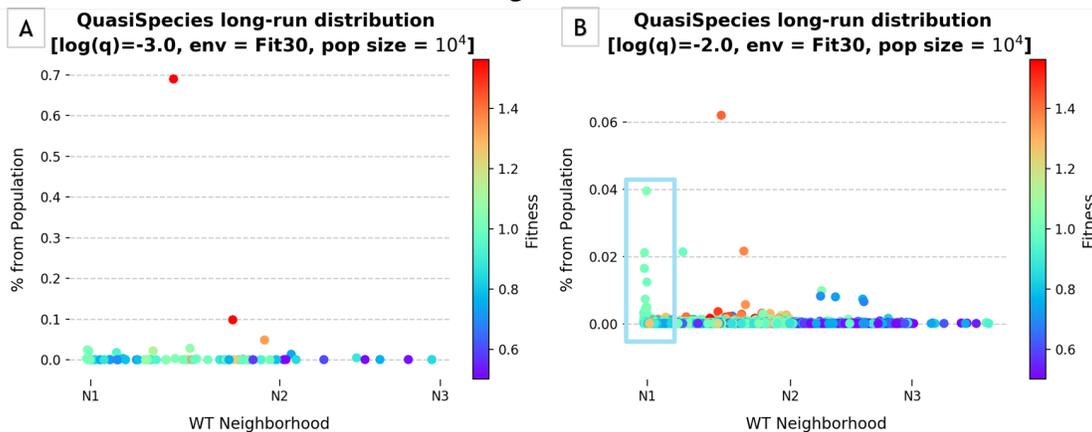
We still didn't manage to bridge this mutation gap, but we do have some clues. First, a paper recently published [37] claims that highly expressed genes and specifically tRNA genes have at least one order of magnitude higher mutation rates. Second, Sardanye's et al. [30] show that q_c is lower when considering spatial effects which we didn't yet. Third, we only simulated each environment condition alone. It might be that using some effective FLS (combination of some sort of the different conditions) will decrease fitness difference between the fittest and the flattest and thus reduce q_c . Finally, the mutation rate could have changed along evolution. Stressful situations tend to increase mutation rate (See for example [33]) and it might be that the WT prevailed thanks to its flatness when such events occurred. Maybe it is better to compare our resulted q_c to some effective mutation rate computed based on assumed mutation rates along evolution.

Figure 29:



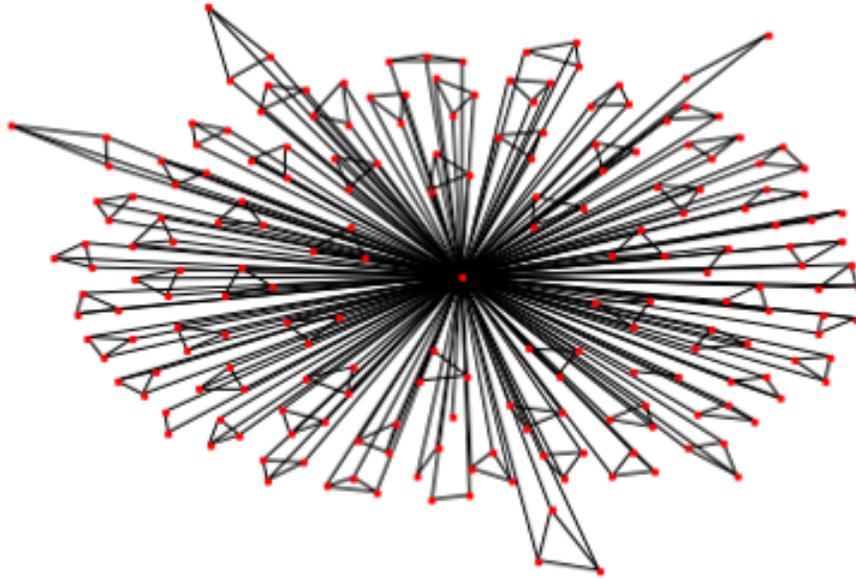
Possible evidence for Survival of the Flattest effect. QS simulation results. A) Population average fitness against $\log_{10}(\text{mutation rate per base pair per generation})$ exhibits a sharp decline in around $q_c \approx 10^{-3}$ from high fitness values to $\bar{f} = 1$. Two conditions are shown. B) Population entropy as a function of $\log_{10}(\text{mutation rate})$. At the same q_c we witness a sharp increase in entropy which indicates a transition in population composition from a few dominant genotypes to more heterogeneous population comprising many different genotypes. C) Fraction of the $N_1^{WT} \cup WT$ group of genotypes out of the entire population as a function of $\log_{10}(\text{mutation rate})$. Again, a sharp increase around q_c which indicate a shift of the population towards the WT and its neighborhood.

Figure 30:



Shift towards the WT peak. Population abundance at the end of two representative runs with different mutation rates. Both panels present abundance as a function of the neighborhood affiliation of each variant relative to the WT (x axis is discrete). Each circle is a genotype, color coded by its fitness. A) Survival of the fittest is still in effect. Approximately 70% of the population is still comprised of one of the high fitness (1.5) genotypes. B) The rise of the QS (see boxed area). Many of the N_1^{WT} variants increase in abundance whereas the fittest decrease. Note the different scales of the y axis.

Figure 31:



Graphical presentation of the WT tRNA gene connectivity. Each dot is a genotype, the central dot is the WT. Lines connect 1-neighbors. The depicted triangles are due to the 4 letter alphabet and the fact that the WT region was densely sampled. It was hard to visualize 2-neighbors, but other Figs such as fig 20 on page 26 or 18 on page 23 which show the WT's neighborhoods in more perspectives can give some intuition.

4.5.3 Is High Connectivity the Culprit?

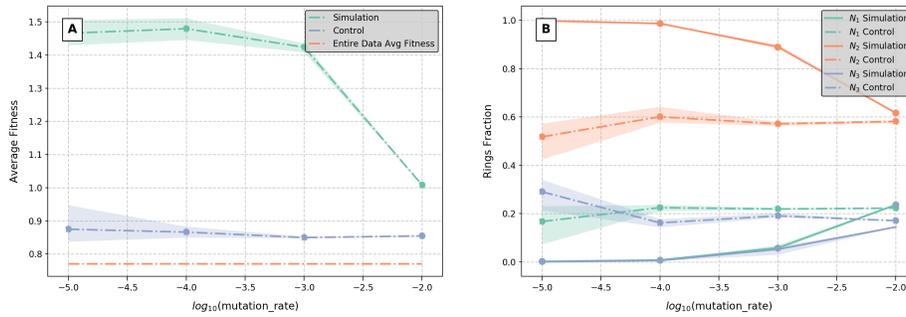
The second concern arises when considering the sub-sampling bias away from the WT. The underlying sequence domain of our tRNA FLS is not symmetric. The WT area was densely sampled compared to much sparser sampling of the outskirts. If we define a connectivity measure for genotype i just by the number of its 1-neighbors present in the data, densely sampled regions will also be densely connected. Since our alphabet consists of 4 letters, there are also many short circles in these highly connected regions (See Fig 31) making it less probable to leave due to mutations, once there. This is of course an unnatural phenomenon - real FLSs are symmetric, while ours don't which might be the reason for the WT's prevalence at high mutation rates.

To ensure that the results aren't due to the underlying sequence domain skewed structure and connectivity, we performed the same simulation as above, assuming equal fitness to all genotypes, thus isolating the underlying structure effect. If there was no sampling bias, we would expect the population to spread equally to all directions and for the population average fitness to be the average of all variant's fitness values, regardless of mutation rate.

Fig. 32A presents the population average fitness for both simulation \bar{f}^s and control \bar{f}^c along with \bar{f} (which is defined as the average fitness of the entire FLS), all of them at 30c. \bar{f}^c was calculated by assigning back the original fitness values to prevailing genotypes only at the end of the simulation run. Each point is again the average of the suitable value after a few hundred generations, among a few repetitions. We ran the control

Figure 32:

Control vs Simulation Results



Controlling for sampling bias towards the WT. A) Average population fitness as a function of $\log_{10}(\text{mutation rate})$. Same simulation results for 30C as in fig 29, together with a control simulation with indifference to fitness to test connectivity influence. Colored regions show first and fourth quartiles. Average fitness at 30c over the entire FLS is also plotted as a reference which highlights the fitness advantage connectivity incurs. For low mutation rates there is no apparent connectivity effect. However, as mutation rate increases, the gap between the control and the simulation is slowly closing. B) WT neighborhood fractions out of total population, for simulation and control, against $\log_{10}(\text{mutation rate})$. Again, colored regions show first and fourth quartiles. Convergence of the simulation values to the control can be viewed as well.

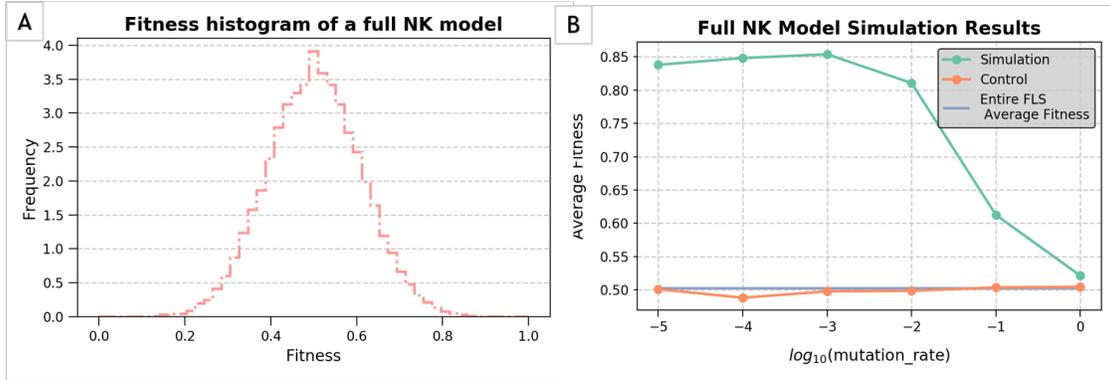
for enough generations to ensure convergence. In a full FLS, assigning all genotypes the same fitness value makes them equivalent. Therefore, with enough time and population size irrespective of the initial condition, all genotypes would be realized and hence the average population fitness would be \bar{f} . We find that $\bar{f}^c > \bar{f}$, meaning that the underlying sequence structure gives a fitness advantage. Interestingly, \bar{f}^c didn't change with mutation rate (which we can regard as a time accelerator). This means that we reached steady state (nothing changes with time), and hence the difference $\bar{f}^c - \bar{f}$ can be considered as the overall underlying structure effect on average fitness.

For low mutation rates there is a clear distinction between \bar{f}^s and \bar{f}^c . However, with mutation rate increasing - each genotype fitness gradually lose meaning in the real simulation while connectivity (and fitness) of the genotype's neighbors gain importance. At $q = 10^{-2}$ only 0.15 units separate the simulation from the control average fitness. When comparing the proportions (out of the entire population) of the first 3 WT neighborhoods (N_1, N_2, N_3) of the control and the simulation (See fig 32B) we see a similar picture.

These results are optimistic for low mutation rate, since there seems to be a clear distinction between the simulation and the control, but inconclusive for higher rates. On the one hand, it seems that the simulation's neighborhood fractions converge to the control which raises a question about conclusions drawn based on Fig 29C on page 36. On the other hand, there is still a significant gap between \bar{f}^s and \bar{f}^c , even at high mutation rates.

To clarify the picture, we decided to go one extra step, doing yet another control. We performed the same simulations as above, only this time on an artificial FLS created using the NK model (see page 55 for more details about the model). The big advantage is that now we can compute an entire FLS and use it as a reference to a partial PCR-like sub sampled FLS. Moreover, since the WT tRNA gene in Chuan et al. study possess both flatness and connectivity, we can also use this model to decouple their effects on the simulation results.

Figure 33:



Experimenting with a NK model full FLS. A) Fitness distribution of a NK model in which we chose $N = 7$, $K = 1$, normally distributed around 0.5. B) QS dynamics simulation results. Average population fitness \bar{f}^s against $\log_{10}(\text{mutation rate})$, compared to a control simulation with equal fitness to all genotypes (\bar{f}^c , control). Population size = 5000, random initialization, ran for a few hundred generations. Each dot is an average of a few replications in each parameter setting. The perfect symmetry in every direction of a full FLS gives no connectivity advantage, such that \bar{f}^c coincide with \bar{f} - the mean fitness over the FLS. For low mutation rates we observe Survival of the Fittest, while the drop later on is due to a threshold crossing (q_c or q_e),

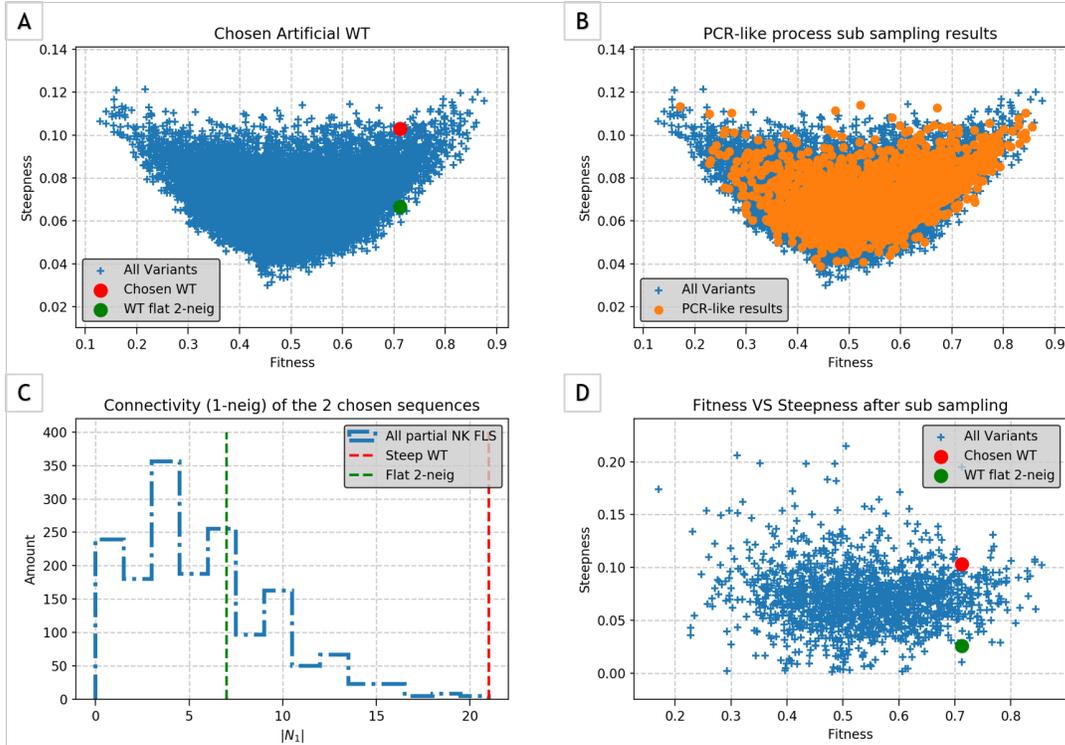
4.5.4 NK Model Simulations, Connectivity and Flatness Decoupling

Simulating QS evolutionary dynamics on the tRNA FLS Chuan et al measured resulted in inconclusive results. On the one hand, it seems that the simulation's neighborhood fractions converge to the control (see Fig. 32 on the previous page) which raises a question about conclusions drawn based on Fig 29C on page 36. On the other, there is still a significant gap between \bar{f}^s and \bar{f}^c , even at high mutation rates. In this part we fabricate a full FLS based on the NK model [12, 11][12, 11] (See page 55 for more details about the model). We then use it to simulate dynamics on different fragments of the FLS, including the entire FLS, in order to better understand simulation results on the tRNA FLS.

To create the NK FLS, we used 7-positions long sequences ($N = 7$), alphabet of size 4 ($|A| = 4$) illustrating nucleotides, and a mild ruggedness coefficient $K = 1$ (Fig 33A). This resulted in a sequence domain containing $4^7 = 16,384$ possible genotypes. Our aim was to first see how evolutionary dynamics unfolds on a full FLS, when all possible sequences are accessible. Thus, we simulated the QS dynamics using a population of 5000 individuals in various mutation rates while tracking average population fitness \bar{f}^s , and \bar{f}^c (which is again a neutral simulation, irrespective of fitness. see Fig. 33B and for comparison Fig. 32 on the previous page). Here, since there is no connectivity difference bias (full FLSs are symmetric), \bar{f}^c and \bar{f} coincide. Still, \bar{f}^s drops sharply at about $q = 10^{-2}$, supporting that the original sharp drop observed in Fig. 29A on page 36 wasn't due to connectivity bias but due to the crossing of a mutation threshold as predicted by the QS theory.

Next, we set out to find proper candidates for WT representation. By decoupling connectivity from flatness, we were hoping to disentangle their effects on the simulation results. Hence, we calculated steepness scores for all variants (Fig 34A), we chose a pair of equi-fitness sequences, 2-neighbor of one another, such that one is relatively flat (denote as x) and the other significantly steeper (denote as y). We then recursively sampled the full NK FLS, where sampling density decreases further away from y , thus mimicking the experimentally measured

Figure 34:



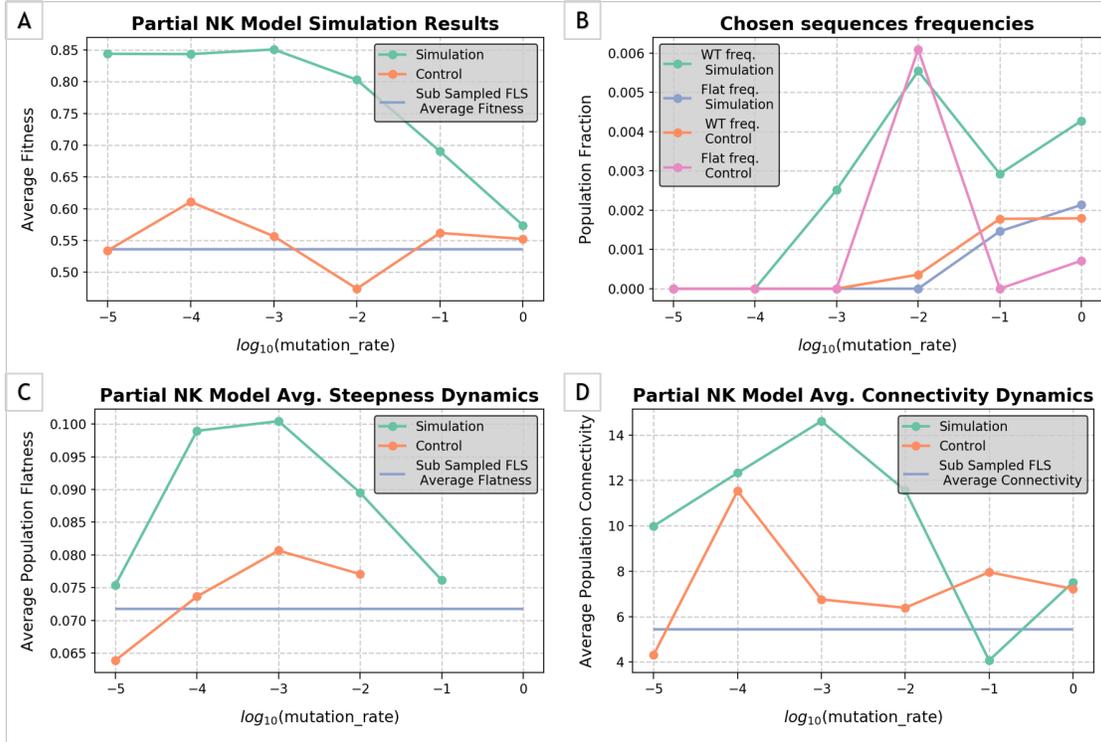
Sampling the full NK FLS. In order to choose a “WT” to be the center of a new FLS sub sampled from the full NK model, we first calculated its steepness to fitness correlation. A) Depicts the correlation, together with a steep sequence y (in red) we chose to be the WT, around it we will sample the rest of the FLS (as Chuan et al. did). We also found a flat 2-neighbor x with equal fitness. We hoped using them will yield connectivity and flatness factors decoupling. B) The results of the sub sampling around y using a simulated PCR-like process. We sampled around 1700 sequences ($\sim 10\%$, out of 16000), reaching almost all parts of the FLS, even though densely sampled around y . C+D) Showing we got our desired traits. y was well connected as opposed to x , while the latter was more flat after recalculating fitness vs steepness correlation of the sub-sampled population.

(using PCR) tRNA FLS. This process resulted in a sample of approx. 1700 sequences around y , x being one of them. Indeed, after the sampling procedure, x remained relatively flat while y had high connectivity, as we wanted (see 34C,D).

The last step was to simulate QS dynamics again, this time on the partial FLS fragment surrounding y , while tracking average fitness \bar{f}^s and \bar{f}^c depend on the mutation rate. We also tracked x and y frequencies, average steepness (defined as $\bar{s} = \sum_{i=1}^N s_i \cdot x_i$, s_i is the steepness score of variant i and x_i is its frequency) and average amount of 1-neighbors (defined as $\bar{N}_1 = \sum_{i=1}^N N_{1,i} \cdot x_i$ where $N_{1,i}$ is the amount of 1-neighbors of variant i) as a connectivity proxy. Unfortunately the results were inconclusive. x and y frequencies were very low (zero in most cases) in almost all mutation values, while Survival of the fittest was prevalent (see Fig on the next page A and B). We couldn't find a pattern looking at \bar{s} and \bar{N}_1 either (see Fig on the following page C and D). However, also here $\bar{f}^c = \bar{f}^s$, signifying no connectivity effect. Overall, we didn't obtain similar simulation results to the tRNA results, let alone use the NK model to better explain them.

We can think of a few reasons responsible for this unsatisfactory outcome. It might be that the connectivity difference between y and the other variants wasn't big enough as in the tRNA FLS. Using sequences of length

Figure 35:



Simulation results for the (PCR-like) NK model sub sampled FLS. Population size = 3000, random initialization, ran for a few thousand generations. Each dot is an average of a few replications in each parameter setting. A) Average population fitness \bar{f}^s , \bar{f}^c against $\log_{10}(\text{mutation rate})$ along with \bar{f} . \bar{f}^c and \bar{f} again seem to agree. \bar{f}^s declines in a similar manner to the dynamics on top the full FLS, albeit less sharply, suggesting it is not due to connectivity bias but rather mutation rate threshold crossing. B) The fraction of the chosen to be WT sequence y (green) and flat reference sequence x (blue) against $\log_{10}(\text{mutation rate})$ shows almost no representation to both of them, even at high mutation rates. Tracking individual sequences might be meaningless in this context. C+D) Average population steepness, defined as $\bar{s} = \sum_{i=1}^N s_i \cdot x_i$, s_i is the steepness score of variant i and x_i is its frequency, against $\log_{10}(\text{mutation rate})$ in C. Average population connectivity, defined as $\bar{N}_1 = \sum_{i=1}^N N_{1,i} \cdot x_i$ where $N_{1,i}$ is the number of 1-neighbors of variant i , against $\log_{10}(\text{mutation rate})$. In both panels, both values decrease after $q = 10^{-3}$ supporting the possibility that the sharp decrease in panel A is due to error threshold.

$O(10^2)$ will make it possible, but since it is practically impossible to store in memory an entire FLS of that size, we had to create a whole new simulation framework that would work on FLS partial sampling. Working with a higher K may also yield better resemblance and outcomes. High ruggedness causes short adaptive steps which can in turn lead to the survival of the less than fit. It is possible that also higher resolution of mutation rate might reveal hidden patterns of the dynamics we are not currently aware of. Lastly, it might be that the NK model fails to capture real FLS properties but other models such as RMF will do better.

To conclude, although attractive, this idea was yet to bear fruits. We witnessed “Survival of the Fittest” and error threshold effect with neither connectivity nor flatness having apparent impact on the simulation results. Still, it equipped us with further ideas (pretty different from each other, currently out of scope) to properly adjust the model to better resemble Chuan et al. FLS. Addressing some of these possible reasons can give rise to interesting conclusions relevant in broader context such as the possible advantages of connectivity bias.

5 Discussion

Using FLS data is a promising framework to better understand evolutionary dynamics. In our work we tried to find the underlying causes that steered evolution towards the *tRNA*^{CCU-Arg} WT. The first staggering result was that the WT is not the fittest in each of the 4 conditions. At least 2.5% of the total variants pool have higher fitness than the WT. We then tried in various ways to check that observation, convincing ourselves it is probably true. Having said that, we are currently considering to perform one of few possible experiments to further confirm that. The first one, which addresses both of our concerns at once, is done simply by synthesizing a large proportion of the high fitness tRNA sequences, including the WT, either transform them into yeast cells or introduce them to $\Delta tRNA^{CCU-Arg}$ yeast via plasmid, and let them grow separately while measuring their OD. This won't be susceptible to background mutations nor to sampling noise. Another option is to use Chuan et al. library itself in various ways, but that would address only sampling noise concern.

Given that there are tRNA sequences with higher fitness than the WT, the results shown in Figs 10 and 11 raise further questions, regarding their biological advantage. A possible analysis addressing that can be to isolate origin-to-destination nucleotide mutations that increase (decrease) fitness the most, and see how it varies with the tRNA position (which ranges from 1 to 72). We already noted a preliminary observation, namely that T (U) rich mutations might provide more flexibility via PTMs after tRNA transcription.

Next, our results indicate that the identity of the WT can possibly be explained by its robustness to mutations (given a bridged mutation gap). Thus being a WT doesn't necessarily mean being the fittest. Genes optimality should be redefined based on a wider set of considerations such as dynamic stability (e.g. in the light of frequent mutations), besides reproduction rate. Moreover, the WT seems to be robust to environment change as well. This is still an open question for us - we are yet to find how it can serve as an evolutionary advantage, except relying on the cost it takes for a yeast cell to frequently build and degrade replication related machinery. Of course, this is only the case if the fitness values measured are a direct consequence of the exponential phase growth rates (i.e. environment steep variant will constantly change its doubling time). This explanation might change if we also consider advantageous lag and yield phases profile as major high fitness contributors.

Overall it seems that the WT optimizes durability to changes, while managing to keep fitness high enough. This suggests that yeast cells might have suffered from frequent changes during evolution that led to that preference. Harsh conditions (e.g. heat or cold shock) lead to stress which in turn increase mutation rate [29, 10]. Such frequent environmental changes can keep mutation rate (or at least effective mutation rate) high while encouraging environmental flatness. In other words, flatness is a desired property for a yeast tRNA gene and therefore the WT possess it.

An alternative, deemed to be less probable, explanation for the WT's flatness has to do with evolutionary dynamics speed. Since it is the slowest on vast almost-planar surfaces of the FLS, thus makes it hard to get out from, it is more probable to find flat genes than genes on other intermediate (i.e. not peaks) states. However, since we showed in section 4.2 on page 26 that most higher fitness variants are probably realized, we tend to dismiss this as the leading explanation.

We are unaware of other works that highlighted mutation robustness (a.k.a genotype flatness) phenomenon in

organisms, let alone eukaryotes, but only in viruses (See for example [3, 39, 30, 15, 25]). That's not accidental - gene flatness is tightly connected to short genome length (equivalently - to very constrained sequences) and high mutation rates. To illustrate that, if 10% of the positions in a viral genome are lethal if mutated, remedy 1% of it by introducing a flat gene can be crucial for survival. In eukaryotes that's much harder to witness - they are much less constrained and thus with much fewer sensitive positions while most of their genes comprise a very small ($< 10^{-4}$) part of the entire genome and therefore can not really remedy this even by using flatness.

Having said that, our work do suggest that evolution can drive tightly constrained eukaryotic genes with high effective mutation rates, as tRNA genes, to be flat. To date, our obtained results are not sufficient to fully prove our hypothesis. To better do that, similar results should be observed in other genes and genomes. We did tried to that with other existing FLS data sets [27, 31], only that they were too sparse (low number of 1-neighbors) and were only measured in one condition (see Appendix B on page 56). But there's still hope - genes characterized by high expression (as tRNA genes) levels might lead the way to more mid-fitness WT genes that use flatness as a mean of addressing high mutation rates [37].

Moreover, tRNA genes tend to create big "families" of duplicated genes in the same genome (sometimes up to hundreds or thousands of duplicates) to adjust and increase expression levels. The different duplicates are usually identical to each other in most positions, thus can be refereed to as a "QS population". In that case, tRNA flatness might be of essence too. For an erroneous gene duplication not to ruin the organism fitness, the mutated duplicated gene needs to contribute to the total organism fitness not much worse than the original gene. In other words, we postulate that high amount of gene duplications can coerce genotype flatness. Hence, genes with many similar copies are other flat candidates.

Finally, various studies reported increasingly smaller benefits of beneficial mutations once genotypes become better adapted (termed diminishing-returns epistasis) in unicellular microbes and single genes [32, 13, 2, 26]. Assuming a WT gene is a result of adaptive trajectories comprised of consecutive beneficial mutations, diminishing-returns epistasis phenomenon is equivalent to (directional) flatness (See 4.3 on page 27). Thus, this serves as another clue for flatness pervasiveness in organisms.

We will conclude by noting the major obstacles FLS sampling bias has took us through. The absence of most of the FLS fitness values (which, of course, is unavoidable) compelled us to derive satisfactory statistical methods to overcome it. However, a FLS sample big enough might pave the way to extrapolation in various ways. In our work we unsuccessfully tried to do that, forecasting a genotype fitness by simply using tRNAscan-SE tool. Other more sophisticated approaches exist (e.g. Machine Learning based) and surely deserve to be studied to further progress FLS research.

6 Materials and Methods

During our work we used various methods to resolve issues along the way. Some of them, the ones we thought are a necessary part of the reading flow, are already integrated into other chapters. This includes a rough assessment of the probability of background mutations on page 14, confidence intervals calculation for high fitness values on page 16 and Genotype flatness calculation in the light of sampling bias towards the WT on page 28.

Of course we had more than that to tackle. Some can be deemed relatively trivial - these can be reviewed and accessed through GitHub code repository (https://github.com/Ylove3/Rugged_FLS_Exploration) for results reproduction along with the data we used. Our overall analysis is code-implemented in an object oriented fashion using python 3.6. Still, other issues were too exhausting to be described in the relevant context, hence we give a detailed description here.

6.1 tRNA Fitness Values Alignment Across Conditions

Fitness values alignment across conditions is imperative to our work. The measured values in each environment were relative to the WT's fitness (which was set to 1, $f_{WT}^m = 1$) in each condition m and thus incomparable between conditions. However, in order to compute effective FLS (see subsection 4.1 on page 24) and environmental flatness (see subsection 4.4 on page 32) we must compare them. If we know how to compare the reference point of each condition (i.e. the WT's fitness), we will be able to align the different values across conditions.

Chuan et al. measured fitness values in each condition using the following pipeline:

1. Measured frequencies F_0 at T_0 , right before competition, by sub sampling and sequencing the variants pool. This was the base line later used for all variants and conditions.
2. They split the original variants pool to the four different conditions, for at least 3 replications each (30C and 23C were replicated 6 times, the rest 3 times).
3. Then they compete each of them for 24 hrs. In between, after 12hrs, they diluted the colonies by a factor of 1/100.
4. For each condition they again measured frequencies F_{24} at T_{24} , just as the competition ended, using the same procedure as in part 1.
5. Next, they calculated how many generations the WT went through, G^m , during the competition using:

$$2^{G^m} = d \frac{g_{24}^m F_{24}^{m,WT}}{g_0 F_0^{WT}}$$

where:

- (a) G^m - number of WT generations in 24 hours in condition m , assuming an exponential growth rate
- (b) d - dilution factor.
- (c) g_t^m - total number of cells at time t (retrieved from cell density (OD) measure) at condition m .
- (d) $R_t^{i,m}$ - # of reads at time t of variant i at condition m .
- (e) $F_t^{i,m}$ - Frequency at time t of variant i at condition m , $F_t^{i,m} = \frac{R_t^{i,m}}{\sum_j R_t^{j,m}}$ (Hence the product $g_0 F_0^{WT}$ is the amount of WT cells at the beginning of the experiment. Note - the denominator above was equal in all conditions).

6. Extracting G yields $G^m = \log_2 \left(d \frac{g_{24}^m F_{24}^{WT,m}}{g_0 F_0^{WT}} \right)$. In their first paper, with 37C as a condition, G was 11.5 generations in 24 hours.

7. Lastly, they calculated per-generation Darwinian fitness of variant i at condition m using the following fold change calculation:

$$Fitness(i_m) = \left(\frac{R_{24}^{i,m} / R_0^i}{R_{24}^{WT,m} / R_0^{WT}} \right)^{1/G^m} \quad (8)$$

If we embrace Chuan et al. assumption when calculating fold change - the competing population is always at exponential growth phase (no log or yield phase, even though we saw it's not entirely the case, see fig 6 on page 13) - Equation 8 can be turned into

$$Fitness(i_m) = \left(\frac{e^{r_{i,m}t}}{e^{r_{WT,m}t}} \right)^{1/t} = \frac{e^{r_{i,m}}}{e^{r_{WT,m}}} = e^{r_{i,m} - r_{WT,m}}$$

if we regard time as a continuous variable, r_i is the growth rate of the i -th genotype. Therefore the growth rate difference between a genotype i and the WT is the logarithm of what Chuan et al. reported as fitness:

$$\ln(Fitness(i_m)) = r_{i,m} - r_{WT,m}$$

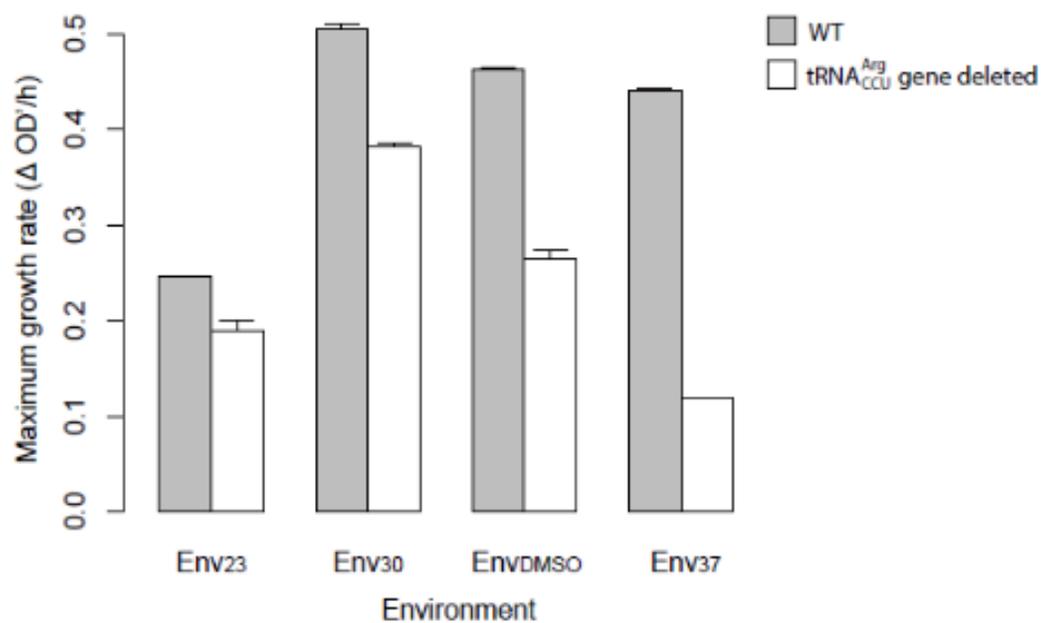
To compare fitness values between two conditions m_1 and m_2 we leave only one of them as a fixed reference, say m_1 , while the other we multiply by $k^{1,2} := \frac{e^{r_{WT,m_2}}}{e^{r_{WT,m_1}}}$ which yields $Fitness(i_{m_2}) \cdot \frac{e^{r_{WT,m_2}}}{e^{r_{WT,m_1}}} = \frac{e^{r_{i,m_2}}}{e^{r_{WT,m_2}}} \cdot \frac{e^{r_{WT,m_2}}}{e^{r_{WT,m_1}}} = \frac{e^{r_{i,m_2}}}{e^{r_{WT,m_1}}}$.

In our case we chose $m_1 = 30C$ to be the across-condition reference. Finally, to calculate the 3 values of k^{m_1, m_i} for the the alignment of the remaining 3 conditions we used Chuan et al. reported WT growth rates (See Fig. 36) in the four different environments.

For example, since the WT growth rate difference between 30C and 23C is approximately 0.25, we have $k^{30C, 23C} = e^{-0.25}$ and thus we divided all fitness values of 23C by $e^{0.25} \approx 1.28$.

Still, there is a possible caveat. Assuming mutations are negligible in a 24 hours time period, exponential evolutionary dynamics are given by the Replicator equation (see equation 2 on page 7). In this regime, a variant prevails only if its fitness is above the population's average fitness at each condition, \bar{f}^m , which might behave totally different in each environment. Possible answer to that are that the WT comprised a significant

Figure 36:



Supplementary Fig. S1. Maximum growth rates of the yeast wild-type strain (gray bars) and $tRNA_{CCU}^{Arg}$ gene deletion strain (white bars) in four different environments. Growth rates are measured by the maximum increase in OD' per hour in mid-log phase. Error bars show one standard error. OD' , converted from OD by the formula $OD'=0.8324 \times OD^3$, is approximately proportional to cell density.

Growth rates of the WT in different conditions. Adopted from Chuan et al.

proportion from the preliminary variant pool at each condition (about 10%) and that $\bar{f}(m, t) = \vec{F}_{t,m} \cdot \vec{f}_{i,m}$ changed by a relatively small amount, ≈ 0.25 (smaller than the fitness std at each condition which is about ≈ 0.35) from the beginning of the competition ($t = 0$) to its end ($t = 24\text{hrs}$) and in a similar way in all conditions (all competitions started with an average population fitness of approximately 0.7 and ended up with it being around 1).

We are well aware that measuring fitness values in more time points than 2 can increase the certainty of such a comparison. However given the existing data and since we can't physically let the variants compete with each other since they reside in different environments, we think it's satisfactory and meaningful.

6.2 QS simulation on the tRNA FLS

We will elaborate on the last procedure here, mainly because it took the longest to implement and with a few lessons learned (mainly computational in nature). We simulated the QS dynamics using 2 methods: Numerically solving the governing dynamical equations for the Steady State and Agent Based Modeling (ABM). The discrete-in-time, matrix form of the QS equations is given by:

$$\vec{x}(t+1) = \frac{\tilde{Q}W\vec{x}(t)}{\bar{\omega}(t)} \quad (9)$$

where the i 'th entry of \vec{x} is just the frequency of the i 'th genotype x_i of i , \tilde{Q} is the mutation matrix $-\tilde{Q}_{ji} = q_{ji}$, $\tilde{Q}_{ii} = Q$ (Quality assurance coefficient), W is a diagonal matrix with $W_{ii} = f_i$ and $\bar{\omega} = \text{diag}(W) \cdot \vec{x}$ is the population average fitness. We assume that the mutation rate between genotypes j and i only depends on the Hamming distance h_{ji} :

$$q_{ji} = (1 - q)^{L - h_{ji}} q^{h_{ji}}$$

is the transition probability (mutation rate) from genotype j to i where L is genome length, h_{ji} is the Hamming distance between genotypes j and i and q is the mutation rate per base pair. Under some mild technical conditions necessary to apply the Perron–Frobenius theory, Eq. 9 has a unique asymptotically stable equilibrium (or Steady State, SS), $\lim_{t \rightarrow \infty} \vec{x}(t) = \vec{x}$, which is the positive eigenvector of the eigenvalue problem (See for example [1] for more details)

$$\tilde{Q}W\vec{x}(t) = \lambda\vec{x}(t)$$

Therefore, when converged, our SS population distribution is \vec{x}_{SS} , $\lambda = \bar{\omega}_{SS} = \text{diag}(W) \cdot \vec{x}_{SS}$. To solve this dynamical system for the SS, we initialized $\vec{x}(t=0)$ (and thus $\bar{\omega}(t=0)$) in a variety of ways (random initialization, lowest fitness, highest fitness etc.) and then iteratively updated $\vec{x}(t+1)$ and $\bar{\omega}(t+1)$ using our fixed pre-computed $\tilde{Q}W$. Convergence was reached when $|\bar{\omega}(t)\vec{x}(t) - \tilde{Q}W\vec{x}(t)| < \varepsilon$ when ε was a predefined threshold.

Note that the Perron-Frobenius theorem (in our case) states that if the sequence domain is connected, i.e. it is possible to reach any sequence from any other sequence with nonzero probability, then $\tilde{Q}W$ has a single stationary distribution and it converges to it regardless of initial condition (The theorem is usually formulated

for stochastic matrices of Markov chains, but FLS sequence domain together with a distance measure is a Markov chain: each sequence is a state and the transition probability between states is determined by their edit distance). For this reason, when we implemented the simulation (in both ways) we made sure to eliminate all sequences with no neighbors (a.k.a. islands, which of course don't exist in nature and are only an artifact of the experiment design) when initializing the population.

To tackle memory and complexity issues we computed the hamming distance matrix H , $H_{ij} = h_{ij}$ up front. Each row of H was calculated using sparse matrix multiplication, resulting in a $\sim 10^8$ entries matrix stored on disk in an h5 file type which allows easy slicing access. At first, the numerical approach looked promising. However later we figured that the system setup caused built-in faults. Mainly, solving for frequencies and low mutation rates caused numerical errors (machine precision is usually bounded by 10^{-16}) which yielded different SS in each run. It was also hard to track the dynamics along the many generations and it took a long time to converge (computationally exhaustive).

On the other hand, ABM, implemented with the mesa python package [20], was much easier. The package is very robust and can be manipulated to different needs (types of agents and simulations). It has built in aggregator functions for data collection along the simulation, together with a set of methods that can enumerate on many different parameter setups and runs. It even has an option for runs multiprocessing (which we didn't use). Finally, it is conveniently equipped with a java-script-based local browser visualizer that can plot simulation progression as it goes.

We randomly initialized a population of 10000 agents, each of them with its own fitness, genotype sequence, hamming distance from the WT and a unique identity. Based on one's fitness we then executed the simulation (written in pseudo-code) on the next page.

Algorithm 1 Quasi Species Simulation

1. Input:
 - (a) $N \leftarrow$ #Population Size
 - (b) $q \leftarrow$ #Mutation Rate per base pair
 - (c) $m \leftarrow$ #The environmental condition we are working in
 - (d) $g \leftarrow$ #Amount of generations to run
2. Randomly initialize the population
3. Do g times:
 - (a) Go through all the population and for each organism i do:
 - i. $r' \leftarrow$ Draw a number uniformly at random from $[0,1]$ (# We normalized the fitness values)
 - ii. If $r' < f_i$:
 - A. $j \leftarrow$ Choose another organism from the population to death
 - B. $r'' \leftarrow$ Draw a number uniformly at random from $[0,1]$
 - C. if $r'' < 1 - (1 - q)^L$: //The chances of at least one mutation to occur while replicating
 - $r''' \leftarrow$ Draw a number uniformly at random from $[0,1]$
 - if $r''' < q$: //The chances of being a double mutant
 - $\tilde{i} \leftarrow$ Choose a 2-neihgbor of i
 - else: $\#(r''' \geq q)$
 - $\tilde{i} \leftarrow$ Choose a 1-neihgbor of i
 - Replace organism j by \tilde{i}
 - D. else: $\#(r'' \geq 1 - (1 - q)^L)$
 - Replace organism j by i
 - (b) Shuffle the organisms order

Two things to note about our implementation:

1. We neglected other types of mutation such as insertions and deletions. The reason we only allow up to 2 mutations simultaneously is because higher order mutations will be very rare in our simulator setup and might lead to numerical errors.
2. It doesn't necessarily stop upon convergence (e.g. when the population structure reaches equilibrium and stops fluctuating), since it took a lot of time to reach it, with no dramatic changes after a few hundreds generations. For our needs, the picture was clear enough already there.

More information regarding simulation's results and the specific parameters used is available in 4.5 and in the code repository.

References

- [1] Alexander S. Bratus, Artem S. Novozhilov, and Yuri S. Semenov. Rigorous mathematical analysis of the quasispecies model: From Manfred Eigen to the recent developments. *arXiv:1712.03855 [q-bio]*, December 2017. arXiv: 1712.03855.
- [2] Hsin-Hung Chou, Hsuan-Chao Chiu, Nigel F. Delaney, Daniel Segrè, and Christopher J. Marx. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. *Science*, 332(6034):1190–1192, June 2011.
- [3] Francisco M. Codoñer, José-Antonio Darós, Ricard V. Solé, and Santiago F. Elena. The Fittest versus the Flattest: Experimental Confirmation of the Quasispecies Effect with Subviral Pathogens. *PLoS Pathogens*, 2(12):e136, 2006.
- [4] J. Arjan G.M. de Visser and Joachim Krug. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, 15(7):480–490, July 2014.
- [5] Inês Fragata, Alexandre Blanckaert, Marco António Dias Louro, David A. Liberles, and Claudia Bank. Evolution in the light of fitness landscape theory. *Trends in Ecology & Evolution*, 34(1):69–82, January 2019.
- [6] Tamar Friedlander, Roshan Prizak, Nicholas H. Barton, and Gašper Tkačik. Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nature Communications*, 8(1), December 2017.
- [7] Sergey Gavrilets. High-Dimensional Fitness Landscapes and Speciation. In Massimo Pigliucci and Gerd B. Müller, editors, *Evolution—the Extended Synthesis*, pages 45–80. The MIT Press, March 2010.
- [8] John H. Gillespie. *Population genetics: a concise guide*. Johns Hopkins University Press, Baltimore, Md, 2nd ed edition, 2004.
- [9] Allan Haldane, Michael Manhart, and Alexandre V. Morozov. Biophysical Fitness Landscapes for Transcription Factor Binding Sites. *PLOS Computational Biology*, 10(7):e1003683, July 2014.
- [10] Erich Heidenreich. Adaptive mutation in *Saccharomyces cerevisiae*. *Critical Reviews in Biochemistry and Molecular Biology*, 42(4):285–311, August 2007.
- [11] Stuart Kauffman and Simon Levin. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128(1):11–45, September 1987.
- [12] Stuart A. Kauffman and Edward D. Weinberger. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of Theoretical Biology*, 141(2):211–245, November 1989.
- [13] Aisha I. Khan, Duy M. Dinh, Dominique Schneider, Richard E. Lenski, and Tim F. Cooper. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science*, 332(6034):1193–1196, June 2011.

- [14] S. Kryazhimskiy, G. Tkacik, and J. B. Plotkin. The dynamics of adaptation on correlated fitness landscapes. *Proceedings of the National Academy of Sciences*, 106(44):18638–18643, November 2009.
- [15] Adam S. Luring, Judith Frydman, and Raul Andino. The role of mutational robustness in RNA virus evolution. *Nature Reviews Microbiology*, 11(5):327–336, May 2013.
- [16] 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.
- [17] Chuan Li, Wenfeng Qian, Calum J. Maclean, and Jianzhi Zhang. The fitness landscape of a tRNA gene. *Science*, 352(6287):837, May 2016.
- [18] Chuan Li and Jianzhi Zhang. Multi-environment fitness landscapes of a tRNA gene. *Nature ecology & evolution*, 2(6):1025–1032, June 2018.
- [19] Michael Lynch, Way Sung, Krystalynne Morris, Nicole Coffey, Christian R. Landry, Erik B. Dopman, W. Joseph Dickinson, Kazufusa Okamoto, Shilpa Kulkarni, Daniel L. Hartl, and W. Kelley Thomas. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 105(27):9272–9277, July 2008.
- [20] David Masad and Jacqueline Kazil. Mesa: An Agent-Based Modeling Framework. page 8, 2015.
- [21] J. Neidhart, I. G. Szendro, and J. Krug. Adaptation in Tunably Rugged Fitness Landscapes: The Rough Mount Fuji Model. *Genetics*, 198(2):699–721, October 2014.
- [22] Johannes Neidhart, Ivan G. Szendro, and Joachim Krug. Exact results for amplitude spectra of fitness landscapes. *Journal of Theoretical Biology*, 332:218–227, September 2013.
- [23] Martin A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press, Cambridge, Mass, first edition edition edition, September 2006.
- [24] Uri Obolski, Yoav Ram, and Lilach Hadany. Key issues review: evolution on rugged adaptive landscapes. *Reports on Progress in Physics*, 81(1):012602, January 2018.
- [25] Samuel Ojosnegros, Celia Perales, Antonio Mas, and Esteban Domingo. Quasispecies as a matter of fact: Viruses and beyond. *Virus Research*, 162(1-2):203–215, December 2011.
- [26] L. Perfeito, A. Sousa, T. Bataillon, and I. Gordo. Rates of Fitness Decline and Rebound Suggest Pervasive Epistasis. *Evolution*, 68(1):150–162, 2014.
- [27] Olga Puchta, Botond Cseke, Hubert Czaja, David Tollervey, Guido Sanguinetti, and Grzegorz Kudla. Network of epistatic interactions within a yeast snoRNA. page 6.
- [28] Hendrik Richter and Andries Engelbrecht, editors. *Recent Advances in the Theory and Application of Fitness Landscapes*, volume 6 of *Emergence, Complexity and Computation*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014.

- [29] Gina P. Rodriguez, Nina V. Romanova, Gaobin Bao, N. Cynthia Rouf, Yoke Wah Kow, and Gray F. Crouse. Mismatch repair-dependent mutagenesis in nondividing cells. *Proceedings of the National Academy of Sciences of the United States of America*, 109(16):6153–6158, April 2012.
- [30] Josep Sardanyés, Santiago F. Elena, and Ricard V. Solé. Simple quasispecies models for the survival-of-the-flattest effect: The role of space. *Journal of Theoretical Biology*, 250(3):560–568, February 2008.
- [31] Karen S. Sarkisyan, Dmitry A. Bolotin, Margarita V. Meer, Dinara R. Usmanova, Alexander S. Mishin, George V. Sharonov, Dmitry N. Ivankov, Nina G. Bozhanova, Mikhail S. Baranov, Onuralp Soylemez, Natalya S. Bogatyreva, Peter K. Vlasov, Evgeny S. Egorov, Maria D. Logacheva, Alexey S. Kondrashov, Dmitry M. Chudakov, Ekaterina V. Putintseva, Ilgar Z. Mamedov, Dan S. Tawfik, Konstantin A. Lukyanov, and Fyodor A. Kondrashov. Local fitness landscape of the green fluorescent protein. *Nature*, 533(7603):397–401, May 2016.
- [32] Martijn F. Schenk, Ivan G. Szendro, Merijn L. M. Salverda, Joachim Krug, and J. Arjan G. M. de Visser. Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene. *Molecular Biology and Evolution*, 30(8):1779–1787, August 2013.
- [33] Erika Shor, Catherine A. Fox, and James R. Broach. The Yeast Environmental Stress Response Regulates Mutagenesis Induced by Proteotoxic Stress. *PLoS Genetics*, 9(8), August 2013.
- [34] Peter F. Stadler. Landscapes and their correlation functions. *Journal of Mathematical Chemistry*, 20(1):1–45, 1996.
- [35] Peter F. Stadler and Santa Fe Institute. Towards a theory of landscapes. In Ramón López-Peña, Henri Waelbroeck, Riccardo Capovilla, Ricardo García-Pelayo, and Federico Zertuche, editors, *Complex Systems and Binary Networks*, volume 461-461, pages 78–163. Springer Berlin Heidelberg, 1995.
- [36] Ivan G Szendro, Martijn F Schenk, Jasper Franke, Joachim Krug, and J Arjan G M de Visser. Quantitative analyses of empirical fitness landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01005, January 2013.
- [37] Bryan P. Thornlow, Josh Hough, Jacquelyn M. Roger, Henry Gong, Todd M. Lowe, and Russell B. Corbett-Detig. Transfer RNA genes experience exceptionally elevated mutation rates. *Proceedings of the National Academy of Sciences*, 115(36):8996–9001, September 2018.
- [38] E. D. Weinberger. Fourier and Taylor series on fitness landscapes. *Biological Cybernetics*, 65(5):321–330, September 1991.
- [39] Claus O Wilke, Jia Lan Wang, Charles Ofria, Richard E Lenski, and Christoph Adami. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *PLoS Biology*, 4(12):3, 2001.
- [40] Sewal Wright. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. volume 1, 1932.

7 Acknowledgments

We would like to thank Dr. Chuan Li and Prof. Zhang for their constant willing to help, explain and share data and conclusions. We also want to mention Dr. Orna Dahan, Dr. Roni Rak, Dvir Shirman and other members of the Weizmann's Pilpel lab for their stimulating talks and ideas, each took this project a step forward. Last, but definitely not least, I would like to thank and appreciate the enormous efforts, time, ideas, encouragement and new ways of thinking Tzachi and Tamar equipped me with all along this tedious and bumpy way.

A Fitness Landscapes as an Optimization Problem

The main interest in FLSs is to study their structure and how it affects the evolutionary process. As mentioned, selection drives populations up the FLS, and the premise is that on FLSs of different structure dynamics would look differently and perhaps even the identity of the WT and the favorite adaptive trajectories leading to it.

To explain a bit further, a more formal definition of a Fitness Landscape, stated as an optimization problem, can be useful. FLS \mathcal{F} is a triplet (S, f, d) . S is the entire possible genome space: $\{a_1, a_2, \dots, a_t\}^N$ where $a_i, 1 \leq i \leq t$, is a possible character in the alphabet of our genome A ($a_i \in A$), N is the length of our genome (equivalent to L in the previous section. Kept here as N for historical reasons). d is the metric between two genomes (usually the Hamming distance) that defines the neighborhood structure and $f : S \rightarrow \mathbb{R}^+$. S can be also viewed as the vertices of a graph, $G(V, E)$ where $V = S$, and $E = \{(s, s'), s, s' \in V \mid d(s, s') = 1\}$ are the edges. Now, one can ask, given a FLS \mathcal{F} , how to find where f is maximal:

$$v^* = \operatorname{argmax}_{v \in S} f(v).$$

The vertex (genome) solving this problem is highly sensitive to the structure of \mathcal{F} . If there is just one peak in the landscape (in other words f is globally concave) then it will be a rather easy hill-climbing. On the other extreme, the landscape can be very rugged, making it hard to dodge local optima. Furthermore, very long genomes (in nature N can range $10^2 - 10^{10}$, depends on the problem) create huge solution space S which will be hard to search in, with exponentially increasing number of possible trajectories to stride on. One model that tries to capture both of these characters is S. A. Kauffman's NK-model [12, 11] which defines a family of FLSs that can be tuned by the parameters N and K . The former determines the length of the genome while the latter the ruggedness of the FLS by introducing fitness dependencies between positions. Where $K = 0$, each position fitness contribution is determined only by its own identity and hence each point mutation change the fitness slightly and predictably. Where $K > 0$ each position fitness contribution is determined by K other position. One point mutation can lead to a great change with low correlation to the original sequence. More precisely, a fitness f of a point $s = (s_1, \dots, s_N)$ is defined as follows:

$$f(s) = \frac{1}{N} \sum_{i=1}^N f_i(s_i, s_{i+1}, \dots, s_{i+K})$$

where the fitness contribution f_i of the i 'th position depends on the values of K (usually chosen at random) other positions of s . For example, if $S = [0, 1]^3$, ($N = 3$) and $K = 1$, in order to calculate the fitness of the following genome $s = (0, 1, 1)$, we will have to know three functions f_1, f_2, f_3 , each assign fitness to positions s_1, s_2, s_3 of the genome respectively. Thus, position s_1 might be dependent on the 3rd position s_3 of s and f_1 might be: $f_1((s_1, s_3) = (0, 0)) = 0.4$, $f_1((0, 1)) = 0.7$, $f_1((1, 0)) = 0.1$, $f_1((1, 1)) = 0.2$. In our case, $s_1 = 0$ and $s_3 = 1$ hence f_1 yields 0.7 which is the contribution of the first position of s to the entire genome fitness.

Other studies tried to classify FLSs using Fourier decomposition, hoping that the decomposition's coefficients will enable to group different FLSs according to similar properties. Looking at FLS \mathcal{F} on a graph $G(V, E)$

where each node has its own fitness assigned to it, allows one to decompose it. To do that we need first to define the graph Laplacian \mathcal{L} . The definition for D-regular graphs, i.e. graphs for which all vertices have the same degree D (which is the case with full FLSs), is $\mathcal{L} = A - DI$, where A is the graph adjacency matrix and I is the identity.

Now, if \mathcal{F} is a landscape on $G(V, E)$, and we denote $\{\varphi_i\}$ as the complete orthonormal set of eigenvectors of the graph Laplacian $-\mathcal{L}$, we call the expansion:

$$f(x) = \sum_{i=1}^{|V|} a_i \varphi_i(x) \quad (10)$$

the Fourier expansion of the landscape with coefficients $\{a_i\}$. Stadler and Weinberger [34, 38] shows that we can use these coefficients to classify FLSs by their ruggedness. they devised an auto correlation function $r(s)$ between different random walks on $G(V, E)$, different only by s steps. The more rugged the FLS is, it is less correlated and thus $r(s)$ will be smaller. By defining normalized amplitudes $A_i = \frac{|a_i|^2}{\sum_{j \neq 0} |a_j|^2}$, they show that

$$r(s) = \sum_{i \neq 0} A_i (1 - \lambda_i/D)^s \quad (11)$$

where λ_i is the eigenvalue corresponding to φ_i eigenvector, assuming that $G(V, E)$ is a regular graph. If we fix the sequence domain $G(V, E)$ and assign different fitness values to it, only the normalized amplitudes A_i will be different. According to Eq. 11 they will also set the FLS ruggedness. Krug et al. [22, 36] show that the normalized amplitude A_1 , corresponding to the second lowest eigenvalue, describes the additive non-epistatic part of the fitness landscape. The sum of the remaining normalized amplitudes, $F_{sum} := \sum_{i > 1} A_i$, describe interactions (epistasis) of increasing order. Fig 37 presents an example we performed for the NK model. It is clear that F_{sum} is increasing with K, while the additive coefficient A_1 decreases with K.

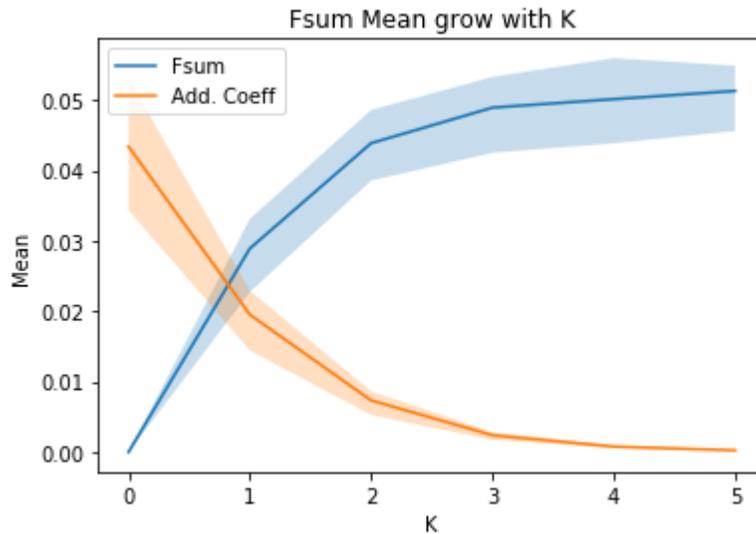
Ruggedness classification is important since the more rugged the landscape is, the tougher it is to climb, reaching the global optimum [12, 11]. Others found different ruggedness proxies [36], although the Fourier decomposition is probably far more informative, albeit available for regular graphs only. Further & more recent analytical works on FLS are available in [28].

B snoRNA and GFP FLSs

As stated in the main text, Chuan et al. weren't the only researches that experimentally measured large scale FLSs. We did consider to use data from two other studies that mapped the local FLS around WT GFP proteins sequence [31] and snoRNA sequence [27]. Unfortunately data sparsity dissuaded us from including them in our analysis. Nevertheless, for the sake of completeness and as a reference, we hereby briefly present the two experiments and a few data snapshots.

In the first experiment Puchta et al. [27] measured the fitness of 60K mutant variants of a 333-long yeast snoRNA in a glucose rich medium at 30C and 37C. The variants were synthesized using error prone PCR and

Figure 37:

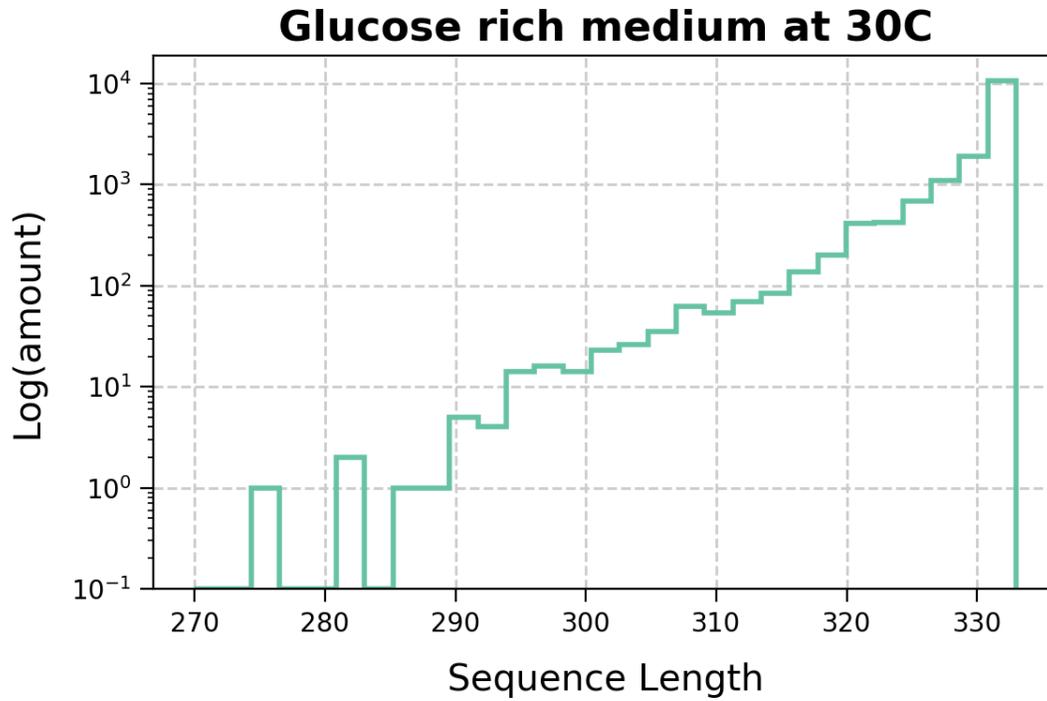


NK model measured ruggedness increases with K. Here we chose $N=6$, $|A|=4$. F_{sum} (blue) and the additive coefficient A_1 (red) against K . Each dot is the mean of several (different) NK model realizations. As expected, the higher F_{sum} is the larger K is, while the additive coefficient is decreasing with K .

later transformed via plasmids into yeast cells, which can grow in galactose-containing medium, but shifting to glucose results in down-regulation of the snoRNA gene and growth arrest. The transformation allows the cells to survive on glucose, but nonfunctional mutants do not support growth. Next, they used sort sequencing relative to the WT (the fitness of the WT was set to be 1) to measure the fitness of each of the variants during competition on glucose. They repeated the competition twice, with high agreement between repeats.

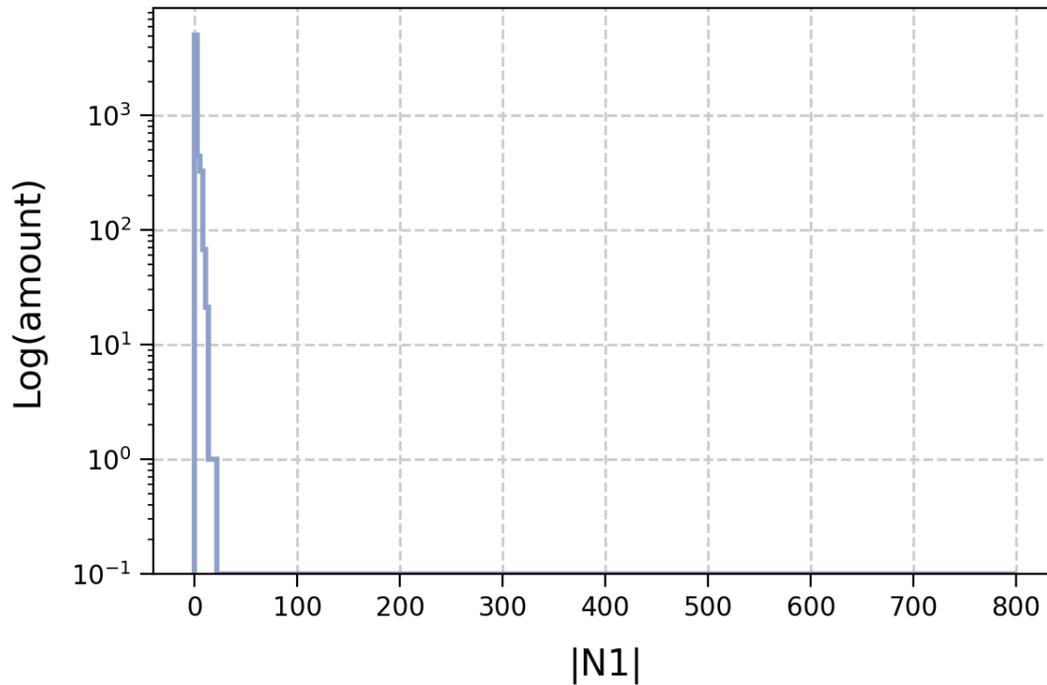
The first major difference from Chuan et al was that many of their mutations were in fact indels as opposed to SNPs (see Fig. 38 showing variants length distribution in one of the repetitions). This made hamming distances computation much harder and hence consisted major obstacle in working with this data set. Using only variants with the same length, after duplicates aggregation, resulted in a very sparse FLS (see $|N_1|$ histogram at Fig. 39). The gene length (~ 4.5 times longer) also contributed to that. Finally, plotting fitness and genotype flatness histograms (see Figs 40 and 41) do show variants with fitness higher than the WT, but in both the WT is not relatively flat. This didn't add an incentive to further investigate the data to our needs.

Figure 38:



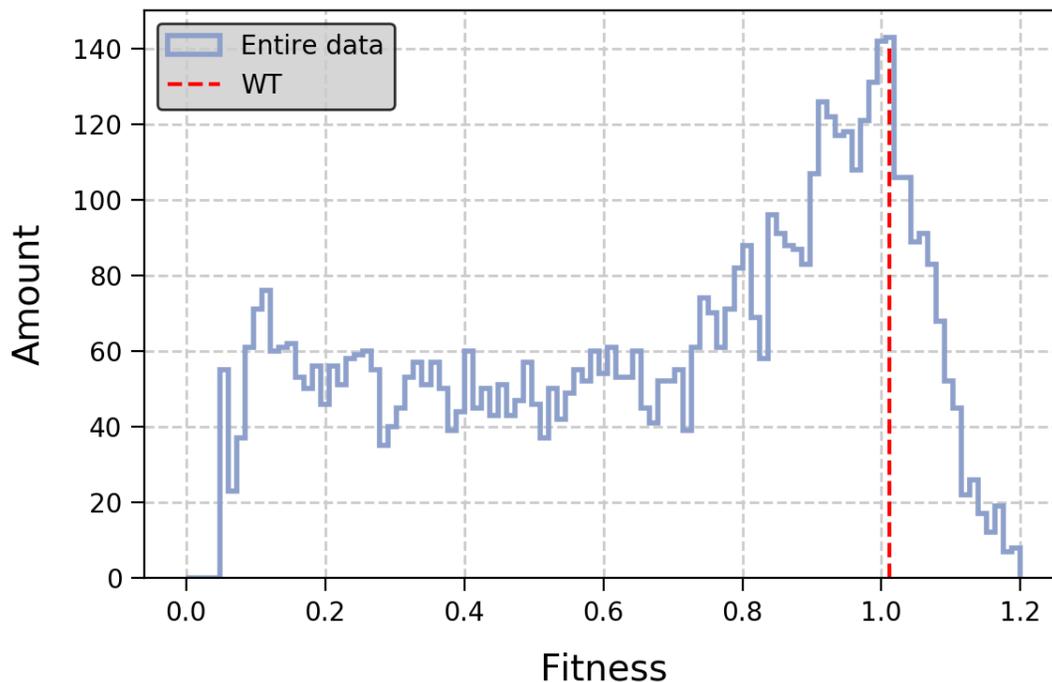
Indels in Puchta et al. experiment [27]. Semi logarithmic histogram of the variants sequence length of one their experimental repetitions. Large proportion of the sequence variants are shorter than the WT sequence which is 333 positions long.

Figure 39:



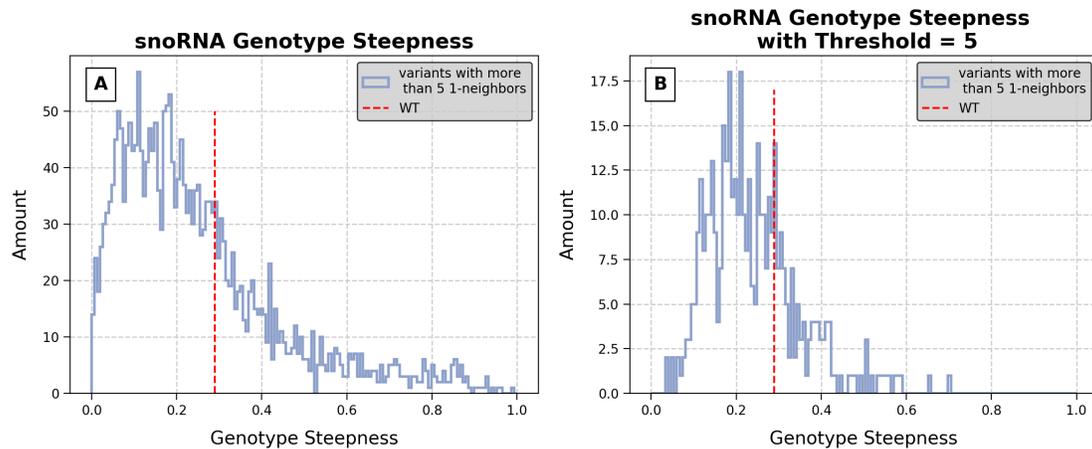
Partially sampled sequence domain, snoRNA FLS. Sampled 1-neighbors histogram for all existing variants of the first replication. Generally, number of k-neighbors for this nt sequence variant equals $3^k \binom{333}{k}$, hence for 1-neighbors it is $3 \cdot 333 = 999$. Unfortunately for most variants less than 2% of their 1-neighbors were measured.

Figure 40:



snoRNA fitness histogram. There are variants with higher fitness than the WT.

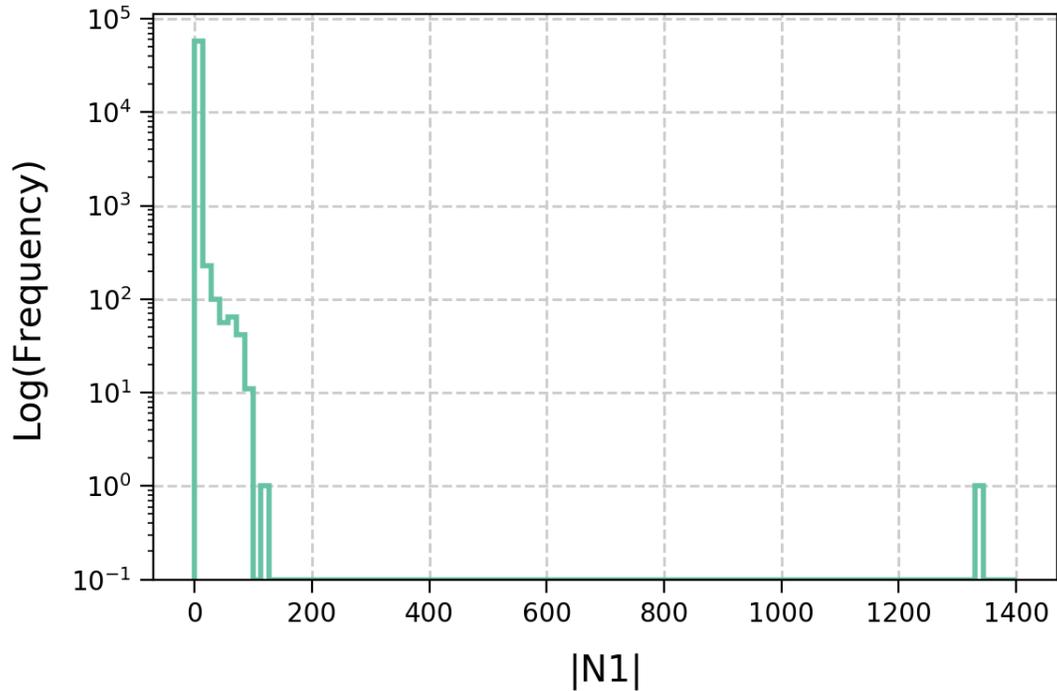
Figure 41:



snoRNA FLS steepness histogram. A) FLS Genotype steepness histogram of 6K variants with the same length. The WT is not relatively flat. B) Same histogram, only for variants with more than 5 neighbors. The WT still seems to be steep compared to other genotypes.

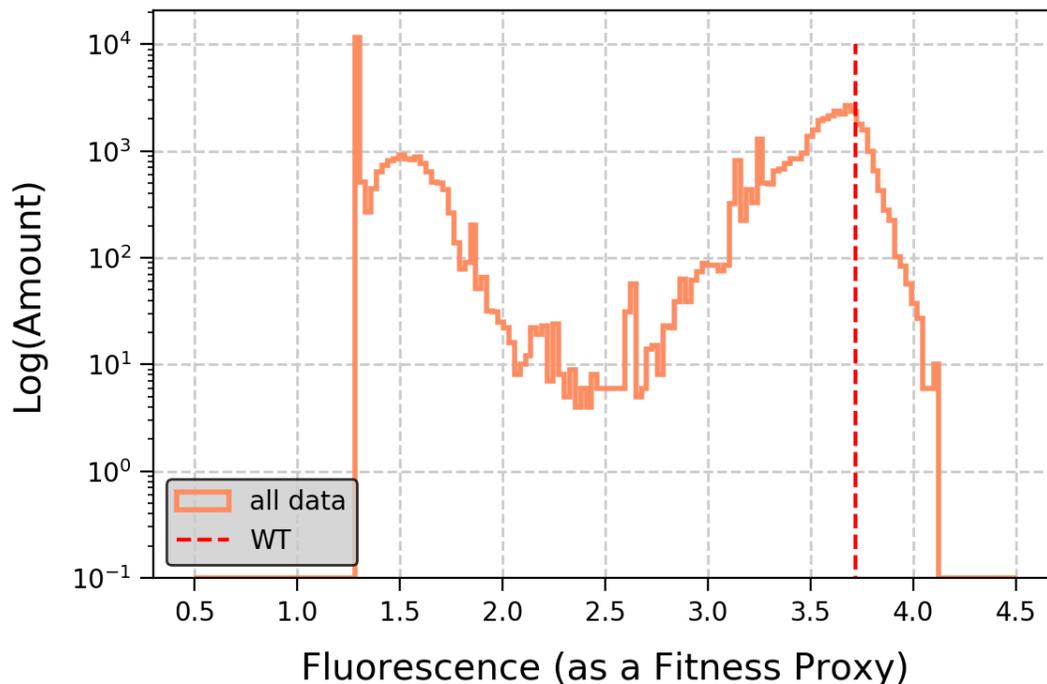
Sarkisyan et al. measured the fitness of 55K mutant variants (50K unique amino acid sequences) of a 705-long *Aequorea victoria* GFP protein, using its native fluorescence phenotype as a fitness proxy. Here the major difference from Chuan et al. is the sequence length (almost ten fold), result in a much larger FLS and thus it is much more susceptible to sparsity (see Fig. 42). Fitness proxy and only one environment measurements are also different. It seems that here as well there are better variants than the WT (see Fig. 43), but large data sparsity is preventing us from concluding if it is due to the WT's flatness or not (Fig. 44).

Figure 42:



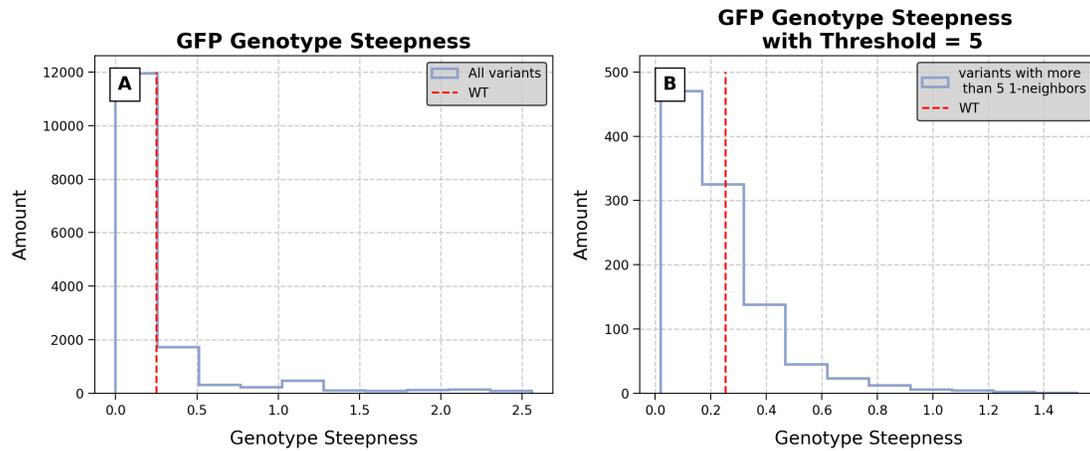
Partially sampled sequence domain, GFP FLS. Sampled 1-neighbors histogram for all existing variants. Generally, number of k-neighbors for this nt sequence variant equals $3^k \binom{705}{k}$, hence for 1-neighbors it is $3 \cdot 705 = 2115$. Unfortunately for most variants less than 0.2% of their 1-neighbors were measured.

Figure 43:



GFP fluorescence (fitness) histogram. There are variants with higher fluorescence than the WT.

Figure 44:



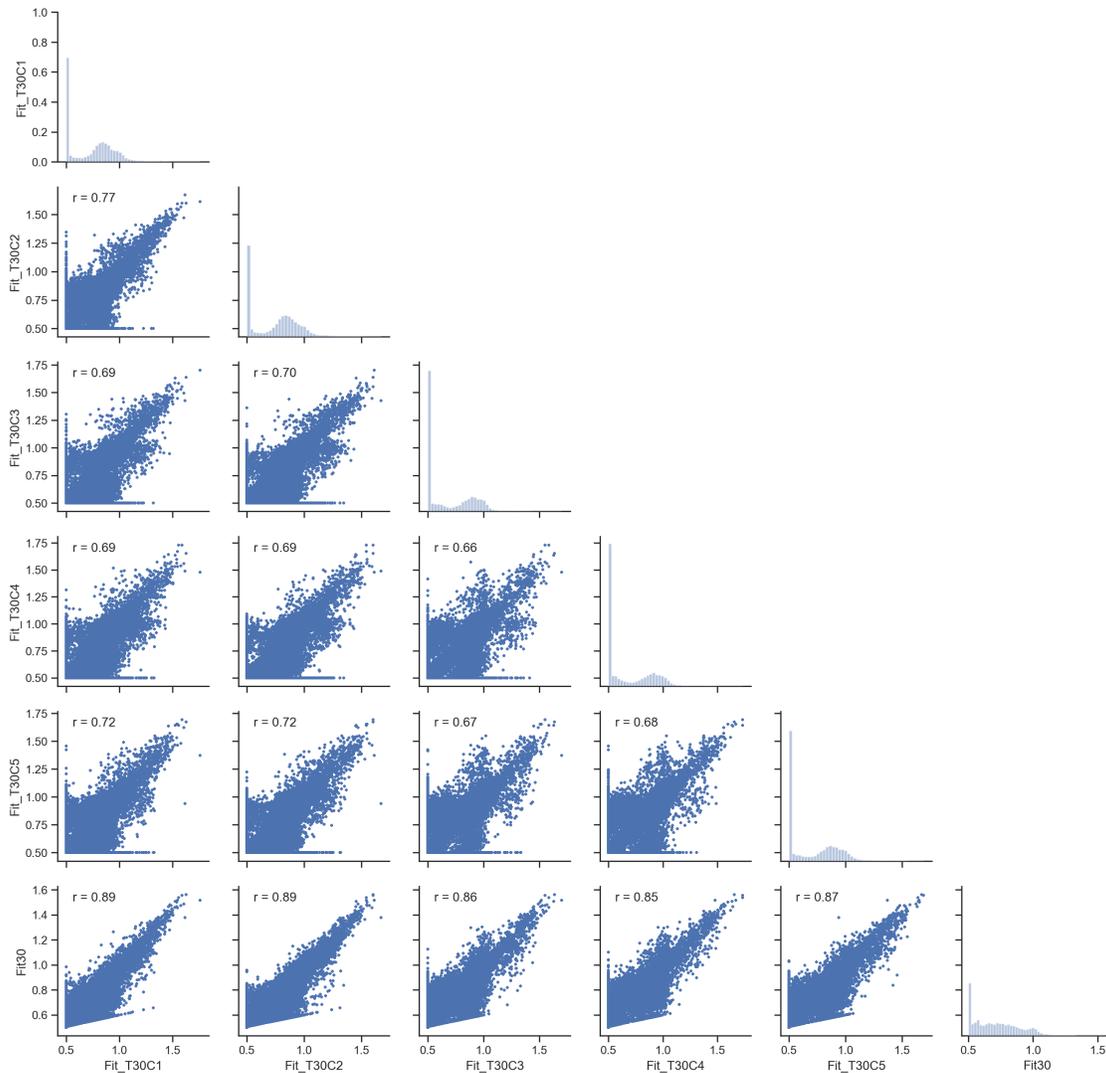
GFP FLS steepness histogram. A) FLS Genotype Steepness histogram of all 60K variants. The WT is not relatively flat. B) Same histogram, only for variants with more than 5 neighbors. The WT still seems to be steep compared to other genotypes. The difference between the two histogram demonstrates the large sparsity effect.

C Fitness correlations between different repetitions

In Fig. 45 through 48 we present the correlations between all the repetitions in all 4 environments Chuan et al used in their experiments. These plots indicate that the error (the disagreement between repeats) is high for low fitness values but low for high fitness values, strengthening the claim that high fitness measurements are not due to measurement errors.

Figure 45:

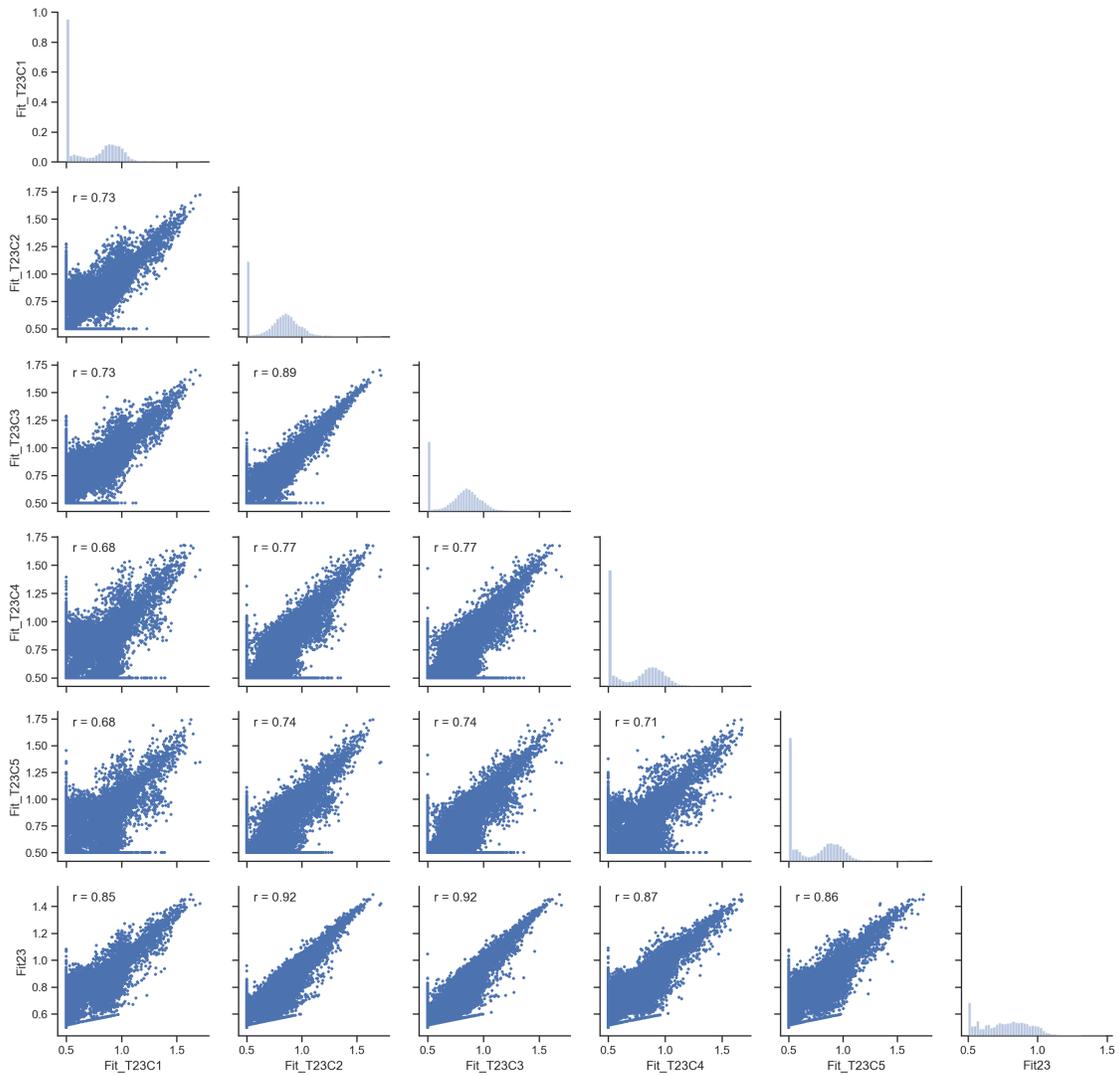
Fitness Values Correlations between Repeats at 30C



Correlation between 6 Repetitions done at 30C. Fitness agreement between repeats in each of the different conditions. On the off-diagonal we present correlation scatter plots and coefficients, including correlation to the final fitness values reported by Chuan et al. (named Fit30). On the diagonal we present the histogram of the fitness values of that specific experiment.

Figure 46:

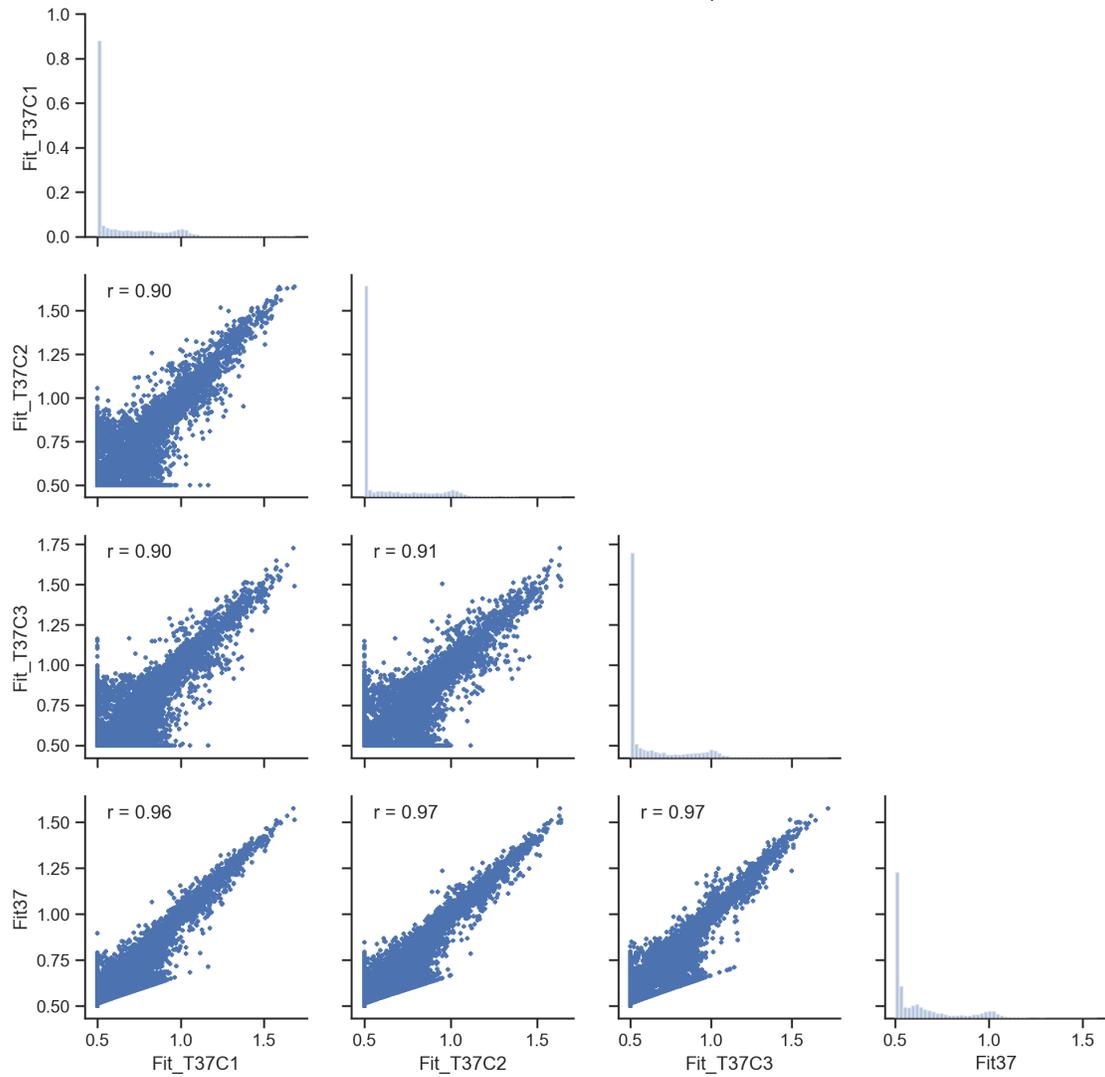
Fitness Values Correlations between Repeats at 23C



Correlation between 6 Repetitions done at 23C.

Figure 47:

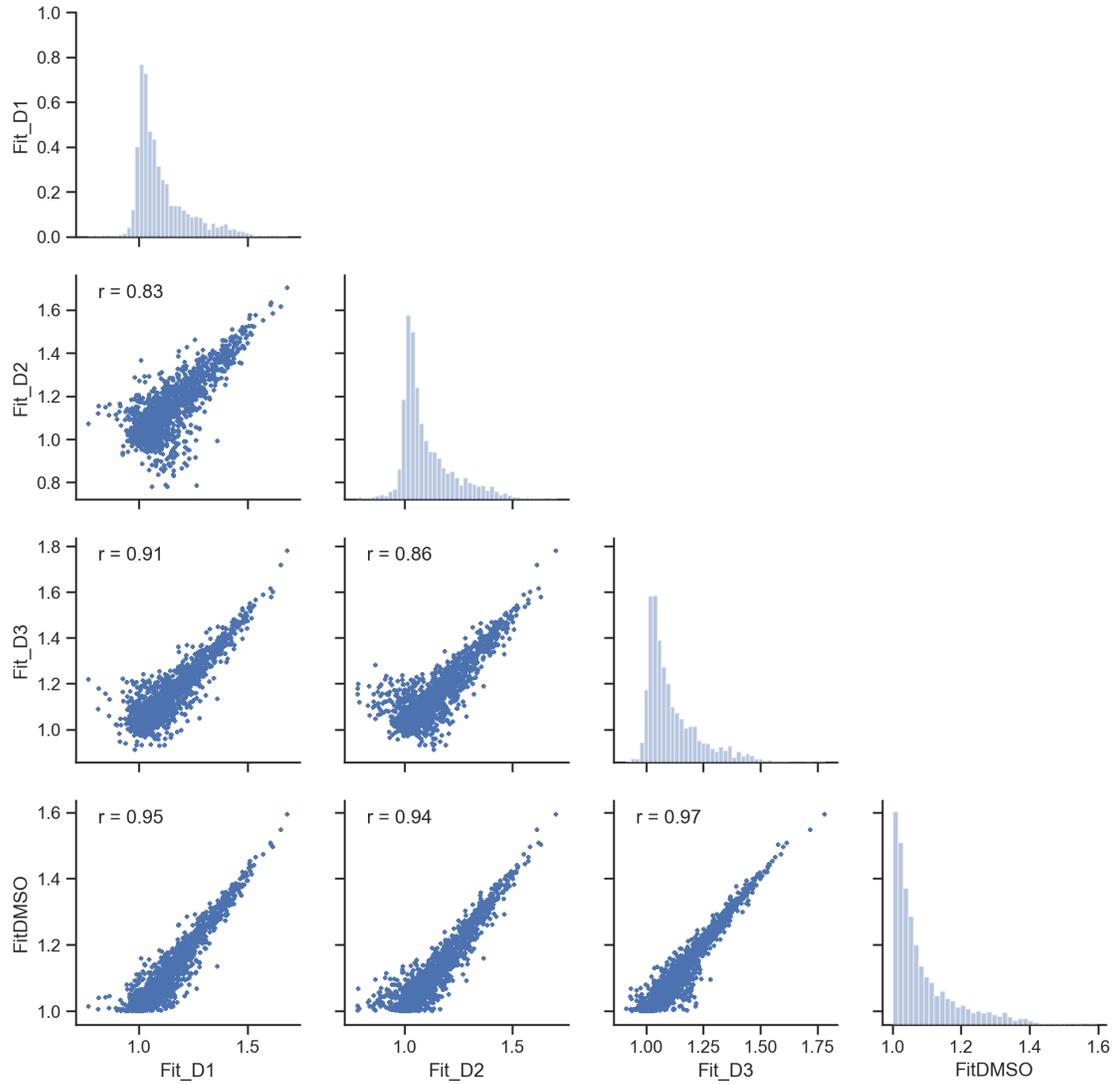
Fitness Values Correlations between Repeats at 37C



Correlation between 3 Repetitions done at 37C.

Figure 48:

Fitness Values Correlations between Repeats at DMSO



Correlation between 3 Repetitions done at DMSO.