The fitness landscape of the codon space across environments

Inês Fragata * [1,+], Sebastian Matuszewski †[2,3,+], Mark A. Schmitz[1], Thomas Bataillon[4], Jeffrey D. Jensen[2,3,5], and Claudia Bank[1]

[1]Instituto Gulbenkian de Ciência, Oeiras, Portugal

[2]School of Life Sciences, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne Switzerland

[3]Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

[4]Bioinformatics Research Centre, Aarhus University, 8000, Denmark

[5]School of Life Sciences, Arizona State University, Tempe, USA

[+]These authors contributed equally

Abstract

Fitness landscapes map the relationship between genotypes and fitness. However, most fitness landscape studies ignore the genetic architecture imposed by the codon table and thereby neglect the potential role of synonymous mutations. To quantify the fitness effects of synonymous mutations, we used a new software based on Bayesian Monte Carlo Markov Chain methods and estimated selection coefficients from deep sequencing data obtained across 9 amino-acid positions from Hsp90 in *Saccharomyces cerevisiae*. This work demonstrates how topology and topography of the codon fitness landscape change when synonymous effects are considered. This impacts how populations traverse fitness space as well as their likelihood of reaching a global optimum, in particular in a stressful environment. Finally, we show that residue position, mRNA stability, and codon frequency are predictors of synonymous effect size. Together these results highlight the role of synonymous mutations in adaptation and demonstrate the potential mis-inference when they are neglected in fitness landscape studies.

*irfragata@gmail.com

†sebastian.matuszewski@hotmail.com

1 Introduction

By considering the relationship between genotype and fitness as a topographic map, Wright (1932) created the concept of a fitness landscape. During the last century this concept has been adopted across various subfields of the sciences, and it has been used extensively to study how populations may adapt to novel environments (Perfeito *et al.*, 2011; De Visser and Krug, 2014; Gorter *et al.*, 2018). Only recently have technological and experimental advances enabled the assessment of large empirical fitness landscapes at high resolution (Weinreich *et al.*, 2006; Hietpas *et al.*, 2013; Bank *et al.*, 2014; Wu *et al.*, 2016; Bank *et al.*, 2016). Wright (1932) noted early on that a complete fitness landscape with L loci, each of which has k alleles, results in a hypercube of k^L genotypes. This enormous dimensionality enforces a careful and limited choice of the mutations that are assayed in any given experiment. Thus, most fitness landscape studies to date have only considered amino-acid changing mutations (e.g. Bank *et al.*, 2016; Wu *et al.*, 2016). This reduction of the genotype-fitness relationship to the amino-acid level poses the danger of misrepresenting the true underlying fitness landscape, and thus the potential routes along which adaptive walks may proceed.

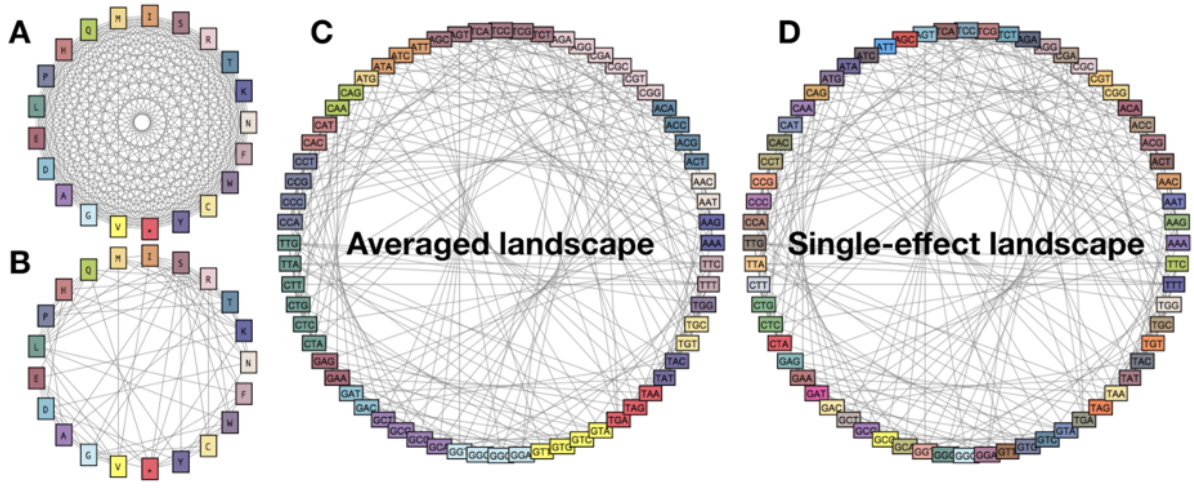

Figure 1: **Codon based and amino-acid based fitness landscapes differ both in topology and topography.** The graphs illustrate potential fitness landscapes at a single amino-acid position. Gray lines indicate single-step mutations and colors indicate potential fitness differences. A) Many studies implicitly assume that all amino-acids are connected by a single mutational step. B) The codon table restricts the number of possible substitutions at the amino-acid level and thus results in a sparser topology. C) Considering the codon level results in a fitness landscape with 64 genotypes. We denote the fitness landscape that neglects the potential effects of synonymous mutations as the averaged landscape. D) We denote the fitness landscape that considers the individual effect of each codon as the single-effect landscape.

Specifically, amino-acid landscapes do not reflect all possible nucleotide mutations present in the genetic code, since they are restricted to 21 genotypes. In particular, even a single amino-acid

position in the genome contains a fitness landscape that consists of the $(4\text{nucleotides})^{3\text{loci}} = 64$ codons at that position. Whereas from the amino-acid view of the landscape, each transition is possible in a single step, the codon based landscape requires up to three mutational steps to transition from one amino-acid to another. This results in a different topology of the fitness landscape with a reduced connectivity (i.e., fewer neighboring genotypes) and larger mutational step size between any two amino-acid genotypes (c.f. Fig. 1 A & B). Moreover, a single-nucleotide mutation in a codon based landscape can result in only 5 to 7 amino-acid changes rather than the 20 total possible amino-acid changes. Thus, at a single amino-acid position, a codon based fitness landscape (with 64 genotypes) can be multi-peaked, whereas the corresponding amino-acid landscape (with 21 genotypes) is by definition single-peaked.

Furthermore, mutations in an amino-acid based fitness landscape are, by definition, non-synonymous. This neglects accumulating evidence from both comparative and experimental studies that synonymous mutations (i.e., mutations that change the codon but not the encoded amino-acid sequence) can display non-negligible fitness effects (Singh *et al.*, 2007; Drummond and Wilke, 2008; Kudla *et al.*, 2009; Zhou *et al.*, 2009; Lind *et al.*, 2010; Plotkin and Kudla, 2011; Sauna and Kimchi-Sarfaty, 2011; Agashe *et al.*, 2013; Bailey *et al.*, 2014; Firnberg *et al.*, 2014; Hunt *et al.*, 2014; Bali and Bebok, 2015; Presnyak *et al.*, 2015; Agashe *et al.*, 2016; Choi and Aquadro, 2016; Knöppel *et al.*, 2016). For example, recent studies have shown that synonymous mutations can affect the speed and accuracy of translation (Drummond and Wilke, 2008; Saunders and Deane, 2010; Plotkin and Kudla, 2011; Bali and Bebok, 2015), mRNA structure (Shabalina *et al.*, 2013; O'Brien *et al.*, 2014; Presnyak *et al.*, 2015), expression in response to environmental changes (Shabalina *et al.*, 2013), and that they are associated with several organismal malfunctions (Parmley and Hurst, 2007; Hunt *et al.*, 2014). Although synonymous effects undoubtedly exist, effect sizes are often small, which has made a systematic characterization difficult. In particular, to our knowledge there exists no study to date that has characterized whether fitness effects of synonymous mutations vary across environments; a finding that could be in concordance with the costs of adaptation that are frequently reported for amino-acid changing mutations (e.g. Bataillon *et al.*, 2011; Wenger *et al.*, 2011; Hietpas *et al.*, 2013; Rodriguez-Verdugo *et al.*, 2014). Thus, when fitness landscapes are defined on the codon level instead of the amino-acid level both its topology (i.e., the number and connectivity of genotypes) and its topography (i.e., the fitness relationship between genotypes; Fig. 1) change. As highlighted by Zagorski *et al.* (2016), a change in the topology of a fitness landscapes can result in dramatically different conclusions about the accessibility of fitness peaks, and the topography further amplifies this effect.

Here we use published data (Bank *et al.*, 2014) from deep mutational scanning (Fowler and Fields, 2014) to study the codon based fitness landscapes of the same 9 amino-acid positions across 6 environments. We first establish that synonymous mutations indeed affect fitness, and then quantify their associated distribution of fitness effects. To this end, we present *empiricIST*, a software that allows for accurate estimation of selection coefficients and credibility intervals from bulk competitions. We then study how considering individual effects of synonymous mutations changes conclusions about both accessibility and ruggedness of the landscapes, and thus the potential for adaptation. By comparing single-effect landscapes to their corresponding averaged landscapes, which neglect the effects of synonymous mutations (Fig. 1 C & D), we quantify how synonymous mutations affect the topography of the fitness landscapes while keeping the topology fixed. Finally, we use regression models to dissect the contribution of environmental *versus* molecular variables to the observed effects of synonymous mutations. Our work provides the first characterization of the distribution of fitness effects of synonymous mutations across environments, and calls for a more careful consideration of synonymous effects in future studies of fitness landscapes and adaptive walks.

2 Material & Methods

2.1 MCMC Method

We provide a software package for 1) processing sequencing count data from deep mutational scanning (DMS) experiments, 2) estimating growth rates using a Bayesian MCMC approach described in detail in (Bank *et al.*, 2014), and 3) post-processing of growth rate estimates to estimate the shape of the beneficial tail of the distribution of fitness effects (DFE). A detailed description of the software, its usage, and options can be found in the accompanying manual (https://github.com/Matu2083/*empiricIST*). In the following, we give a brief description of the assumed experimental setup and the model underlying the MCMC and estimation procedure, and by means of simulations compare the accuracy of the results to that obtained from conventional linear regression (Matuszewski *et al.*, 2016).

Assumptions of the model and input data

We consider an experiment assessing the fitness of K mutants, labelled $i \in \{1, \cdots, K\}$. Each mutant i is assumed to be present at initial population size c_i and to grow exponentially at

constant rate r_i, such that its true abundance at time t, $N_i(t)$, is given by $N_i(t) = c_i \exp^{r_i t}$. At each sampling time point $t \in \{1, \cdots, T\}$, sequencing reads $n_{i,t}$ are drawn from a multinomial distribution with parameters $n_t = \sum_{i=1}^{K} n_{i,t}$ (i.e., the total number of sequencing reads) and $p_t = (p_{1,t}, \cdots, p_{K,t})$, where $p_{i,t} = \frac{c_i exp^{r_i t}}{\sum_{i=1}^{K} c_i exp^{r_i t}}$ is the relative frequency of mutant i in the population at time t. Here, time is measured in hours to make results comparable across different environmental conditions (Chevin, 2011; Bank *et al.*, 2014). The software allows for input of either generation or standard time. We furthermore assume that sampling points are independent such that the overall likelihood can be written as the product of the individual likelihoods of each sampling point.

$$L(n) = \prod_{t \in T} L(c, r | \{n_{1,t}, \cdots, n_{K,t}\}).$$

All initial population sizes c_i and growth rates r_i are estimated relative to those of a chosen reference mutant with its initial population size and growth rate arbitrarily set to 10 000 and 1, respectively. Here, the wild-type sequence in laboratory conditions of $30°C$ was used as the reference.

MCMC model

We implemented a Metropolis-Hastings algorithm in C++ using flat priors allowing all attainable values $r_i \in R^+$ and $c_i \in N$ to be realized with equal probability. During the burn-in period the variance of both proposal distributions was adjusted such that the targeted acceptance ratio is around 25%, which optimizes the performance the MCMC chain (Gelman *et al.*, 1996).

The updated variance of the proposal distribution is calculated using

$$\sigma_{\text{new}} = \sigma_{\text{old}} f(k; y, k)$$

with

$$f(x; y, k) = \left[1 + \frac{(\cosh(x - y) - 1))(k - 1)}{\cosh(y - |x - y|) - 1}\right] \text{sgn}(x - y),$$

where x denotes the targeted acceptance ratio, y is the current acceptance ratio, and k is a (fixed) scale parameter that restricts the maximal change in the variance of the proposal distribution. After discarding the first 100 000 accepted samples (i.e., after the burn-in period), the MCMC

was run for an additional 10 000 000 accepted samples. Only every 1000th sample was retained for further analyses, such that the posterior distribution of each parameter was characterized by 10 000 samples overall.

Convergence and mixing were checked by visual inspection of the resulting trace files for all estimated parameters, and by calculating the effective sample sizes (i.e., the number of independent samples) and the Hellinger distance (Boone *et al.*, 2014) between sets of 1000 batched recorded samples. Effective sample sizes were generally larger than 1000 for all parameters, and Hellinger distances below 0.1 indicated convergence and good mixing. To facilitate estimation, we took advantage of the fact that the multinomial distribution is preserved when a subset of the counting variables are observed. This enabled us to split the data set into sub-data sets with 10 mutants each (implicitly treating the other mutants' sequencing reads as observed). More options such as outlier detection, data imputation, DFE tail-shape estimation are detailed in the Supporting Information.

Assessing accuracy of the MCMC

To assess the accuracy of the Bayesian MCMC approach, we compared its parameter estimates to those obtained using ordinary least squares (OLS) linear regression of the log-ratios against the number of sequencing reads $n_{i,t}$ over the different sampling time points (Matuszewski *et al.*, 2016). For that we simulated time-sampled deep sequencing data (implemented in C++; available from https://github.com/Matu2083/empiricIST), assuming that individual mutant growth rates and initial population sizes for each of the K mutants are drawn independently from a normal distribution (i.e., $r_i \sim \mathcal{N}(1, 0.01)$) and a log-normal distribution (i.e., $c_i \sim 10^{\mathcal{N}(4,0.25)}$), respectively. Without loss of generality, we denote the wild-type reference (or any other reference genotype) by $i = 1$ and set its growth rate to 1. Sequencing reads were then drawn independently for each of the T equally spaced time points from a multinomial distribution with parameters n_t (i.e., the number of total sequencing reads per time point) and $p_t = (p_{1,t}, \cdots, p_{K,t})$. To check the robustness of these results when applied to the real experimental data, we furthermore drew growth rates from a mixture distribution

$$r_i \sim \begin{cases} |N(1, \hat{\sigma})| \text{ if z} = 0, \\ \exp(\hat{\lambda}) + 1 \text{ if z} = 1, \end{cases}$$

where $Z \sim \mathcal{B}(x)$ is a Bernoulli-distributed random variable that indicates whether growth rates are drawn from the deleterious part of the DFE (i.e., if $z = 0$) or from the exponential beneficial

tail (i.e., if $z = 1$). The parameters $\hat{\sigma}$, $\hat{\lambda}$, and $\hat{x}$ are estimated from the underlying experimental data, and based on growth rate estimates obtained from OLS linear regression.

Finally, the accuracy of the parameter estimates was assessed by computing the mean squared error (MSE)

$$\text{MSE} = \frac{1}{K-1} \sum_{i=2}^{K} (\hat{r}_i - r_i)^2,$$

the length of the credibility interval (CI, calculated from the MCMC posterior distribution), and the frequency of the true growth rate lying in the 95% confidence interval calculated over 100 simulated data sets.

Outlier detection in empiricIST

Apart from its main program – the Bayesian MCMC program – *empiricIST* provides Python and shell scripts for data pre- and post-processing. Details about their usage and options are given in the accompanying manual. Here we outline the two different options that are available for dealing with outliers in the sequencing data – i.e., outlier detection and data imputation – and explain the DFE tail-shape estimation (see Outlier Detection in *empiricIST* in SI).

As an alternative to treating outliers as unobserved (i.e., missing data), we also implemented an approach in which data points identified as outliers were imputed (see SI). For that we again used the linear regression of the log ratios of the mutant's read number to the total number of reads at each individual time point (i.e., the 'total' normalization, *sensu* Bank *et al.*, 2014), and classified as outliers data points that exceed the DFBETA cutoff of 2 and that had an absolute studentized residual bigger than 3 . In comparison to other reasonable and established outlier criteria, this approach proved to be more cautious as exemplified by the higher specificity and lower sensitivity (Fig. 2, Fig. S1). By combining two independent outlier criteria (i.e., the DFBETA statistic and the studentized residuals), this approach ensures that data points identified as outliers have leverage effects (i.e., change the slope considerably) and are in conflict (meaning that are very different in comparison) with the remaining data points. Thus, to minimize changes in the original experimental data we took an extremely conservative approach, such that only those data points that stand out as extreme outliers will be imputed.

When comparing the mean squared error (MSE) over 100 simulated data sets across different outlier detection methods, we find that the MSE increases with the proportion of outliers in the data set, independent of the method used. Imputing data points generally improves the accuracy

of the parameter estimates compared to treating outliers as missing data (Bank *et al.*, 2014, ;Fig. S2, S3). Expectedly, when there are no outliers in the data, the wild-type normalization displays the lowest error. However, with only 1% outliers in the data, the error of the wt-normalization is comparable to that of the total normalization and becomes increasingly worse as the proportion of outliers in the data increases (Fig. S2). Note that in the presence of outliers, using any outlier method improves growth rate estimates considerably.

Estimating the shape for the beneficial tail with empiricIST

Finally, *empiricIST* contains a Python script for estimating the shape of the beneficial tail of the DFE. Currently, it is believed that these effects typically follow an exponential distribution (Gillespie, 1983, 1984) characterized by many small, nearly neutral mutations and a few strongly beneficial mutations. Using extreme value theory, it is however possible to test whether experimental data complies with that assumption (and falls into the Gumbel domain), or whether the data is better represented by distributions from the Weibull domain (i.e., distributions that decay more rapidly as an exponential distribution, implying more small-effect mutations) or from the Fréchet domain (i.e., distributions decaying less rapidly than an exponential distribution implying an excess of large-effect mutations; see also Beisel *et al.*, 2007). Additional information about the different types of distributions and likelihood estimation are available in the section on DFE estimation in the SI.

We analyzed the power of the maximum-likelihood method to make this distinction by simulating 1000 Generalized Pareto Distribution (GPD) data sets for different underlying shape parameter (κ) values (spanning across all three GDP domains) and varying sample sizes. We find that for small sample sizes (Fig. S4A, B) $\hat{\kappa}$ displays a large variance and a slight negative bias, in particular, if the underlying shape parameter is from the Weibull domain (i.e., $\kappa < 0$). This bias is caused by a (numerical) discontinuity in the log-likelihood function around κ= -1 (eq. S3 in SI), causing κ to consistently deviate (Rokyta *et al.*, 2008). As sample size increases, however, the variance of the maximum-likelihood estimate decreases and its bias vanishes (Fig. S4C, D). Furthermore, while κ typically falls into the correct domain (even for low sample sizes), the statistical power for detecting deviations from the null hypothesis (i.e., whether H_0: κ=0) is low (unless sample sizes are large).

2.2 Experimental data

The data used in this study were originally obtained in Bank *et al.* (2014) using the EMPIRIC approach (Hietpas *et al.*, 2011, 2012). In this study, all 576 possible single-codon mutations in a 9 amino-acid region of the C terminal part of Hsp90 in *Saccharomyces cerevisiae* were generated and bulk competitions were performed under six different environmental conditions (25°C, 30°C, 36°C, 25°C + S, 30°C + S, and 36°C + S, where S represents the addition of 0.5M sodium chloride). For simplicity, we will refer to these conditions as normal medium or high salt medium, and abbreviate these by 25N and 25S, for example, when additionally referring to the 25°C environment. The experiment was replicated 3 times for the 30N environment and 2 times for the 30S environment. Populations were originally adapted to the 30N environment, thus changes to other environments correspond to shifts from the optimum (Bank *et al.*, 2014).

Growth rates for all mutants were estimated using *empiricIST*. Furthermore, to obtain growth rate estimates per amino-acid (residue) position, we pooled nucleotide sequences and jointly estimated growth rates for those nucleotide sequences that resulted in the same amino-acid sequence (see above and SI). Our downstream analyses are based on 1000 subsamples of the posterior distribution obtained from *empiricIST*, if not otherwise indicated.Selection coefficients were obtained by normalizing to the median growth rate of all mutations synonymous to the reference sequence as detailed in Bank *et al.* (2014).

Distribution of synonymous mutations

We obtained the distribution of synonymous fitness effects (DSE) across all amino-acid mutations as the difference between each individual codon selection coefficient and its corresponding pooled amino-acid estimate. These data were used to perform the analyses in the section on potential mechanisms underlying the effect of synonymous mutations on fitness.

2.3 Detecting the effect of synonymous mutations

Analyses and results to assess experimental error and reproducibility of measurements

To assess the reproducibility of measurements, we compared the correlation between selection coefficient estimates across the three 30N and two 30S replicates, and computed the overlap in their growth rate posteriors. For each replicate pair, we calculated the correlation between mutation-specific fitness effects from both the median estimates and 1000 randomly selected

posterior samples. The median correlation of fitness effects across pairs of replicates was 0.84 (lower and upper credibility intervals from 1000 posterior samples: [0.78, 0.88]) for high salt medium and was 0.98 (lower and upper credibility intervals from 1000 posterior samples: [0.97, 0.99]) in standard medium (Fig. S5) confirming that the experimental protocol has an excellent resolution for measuring selection coefficients. An ANOVA test indicated that experimental error was negligible in comparison to the effect of changing medium (Table S1, Fig. S6) and confirmed the previously observed strong costs of adaptation (Hietpas *et al.*, 2013).

To quantify whether the *empiricIST* credibility intervals cover the experimental error appropriately, we estimated the overlap between the 95% credibility intervals of the posterior distribution for all pairs of replicates. We observed a large overlap between pairs of replicates (Fig. S7, normal environment – a) Rep1-2: 98%; b) Rep1-3: 91%; c) Rep2-3: 90%; high salt environment – d) Rep1-2: 90%), indicating that the variance between replicates is indeed mostly covered by variance in the posterior distribution, and that we can use *empiricIST* credibility intervals as confidence levels in our analysis.

We used linear models to extract the contribution of various factors to the estimated effects of synonymous mutations. Model variable names are highlighted throughout the paper using Italics. The following analyses were performed on the distribution of synonymous effects data, i.e., the data in which the median amino-acid effect was removed.

We estimated the relative contributions of the experimental error and the effect of synonymous mutations in the data by means of three approaches. First, we compared the impact of *replicate*, *codon* and *medium* (i.e., whether salt was added or not) using the following ANOVA model with data between replicates 2 and 3 of both the standard and the high salinity environment for 30C:

$$\begin{aligned} \mathrm{Y} \quad &= \mathrm{codon} + \mathrm{replicate} + \mathrm{medium} + \mathrm{replicate{*}codon} \; + \mathrm{codon{*}medium} \\ &+ \mathrm{replicate{*}medium} + \mathrm{codon{*}medium{*}replicate} \end{aligned}$$

where Y corresponds to the normalized selection coefficient, *codon* to a fixed factor corresponding to the 64 codons present in the data, *replicate* to a fixed factor pertaining to the arbitrary replicate number 2 or 3 for each environment, *medium* is a fixed factor corresponding to the presence or absence of high salt concentration in the medium and ϵ corresponds to the residual error. We estimated effect size by calculating η^2 (i.e., the ratio of the variance explained by a predictor to the total variance explained by the entire model - (Levine and Hullett, 2002)) for each of the model terms, using the etasq function of the *R* package sjstats (Lüdecke, 2017). To

assess the variability of our estimates, we performed the analysis for 1000 posterior samples. Finally, using the distribution of synonymous mutations referred above, we tested the overlap of beneficial mutations across replicates for normal and high salinity environments. For that, we extracted the 30 most beneficial synonymous mutations (approximately corresponding to the 5% beneficial tail) for each replicate, and estimated the overlap across the three 30N replicates, and two 30S replicates.

Quantifying the effect size of synonymous mutations

To quantify the effect size of codon changes, we performed a linear regression for each amino-acid (including all amino-acids with 3 or more codons) and calculated η^2 (as proxy for effect size (Levine and Hullett, 2002)) for the *codon* term. The regression per amino-acid was performed within each environment and took into account *residual depth* (i.e., whether the position was buried or exposed). Pooling of positions was done to allow for the testing of codon effect within amino-acid. To minimize potential differences arising from pooling positions, we separated the data into buried and exposed positions according to residue depth. Additionally, using an ANOVA model we tested how the estimated effect size per amino-acid (using η^2 as dependent variable) varied across *environment* and *amino-acid*.

Finally, we calculated 10 000 pairwise differences between synonymous mutations, between random amino-acid pairs and between random pairs of samples of the posterior to assess the effect of synonymous mutations in comparison with amino-acid changes and in comparison to the variation between posterior samples.

Investigating the effect of synonymous mutations on the topography and the dynamics of adaptive walks in codon fitness landscapes

To quantify the impact of effects of synonymous mutations coding for the same amino-acid on the topography of the fitness landscape, we compared the single-effect landscape with the averaged landscape. For the single-effect landscapes (Fig. 1D) the effect of each codon was directly obtained from the experimental data. For the averaged landscape (Fig. 1C) we assigned to every codon that coded for the same amino-acid the same pooled amino-acid estimate obtained from *empiricIST*.

Each amino-acid position in our data set corresponds to a complete multi-allelic fitness landscape with $3^4 = 64$ genotypes. We characterized the resulting $9 \cdot 6 = 54$ fitness landscapes using several fitness landscape statistics. We estimated 1) the roughness-to-slope ratio (Aita *et al.*, 2001;

Szendro *et al.*, 2013; Bank *et al.*, 2016) to quantify the relative deviations from an additive model; 2) the multi-allelic gamma statistics (Bank *et al.*, 2016; Ferretti *et al.*, 2016) to characterizethe prevalence and type of epistasis in the landscape; 3) the number of local peaks (Szendro *et al.*, 2013); and 4) the length and variance in the length of potential adaptive walks in the landscapes (Neidhart and Krug, 2011; Szendro *et al.*, 2013).

Credibility of the estimates was assessed by computing the fitness landscape statistics for 100 posterior samples.

Potential mechanisms underlying the effect of synonymous mutations on fitness

There are several mechanisms through which synonymous mutations can affect protein translation (reviewed in Plotkin and Kudla, 2011). In this study we focused on whether codon usage frequency or predicted mRNA stability (using as proxy Gibbs free energy and melting temperature) can predict effects of synonymous mutations(Presnyak *et al.*, 2015).

Firstly, to enable the inclusion of codon frequency patterns in yeast into our regression models, we obtained the relative abundance of each codon in the yeast genome from the Codon Usage Database (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4932).
Secondly, synonymous mutations may affect translation through different stability of the mRNA generated by different codons. To obtain predictions of how mRNA stability is affected by synonymous mutations, we ran the prediction software *mfold* (Zuker *et al.*, 1999; Markham and Zuker, 2008), for 25°C, 30°C and 36°C and with high salt concentrations (0.5M Na^+), with physiological concentrations of salt (0. 015M Na^+), and 0.001 M Mg^{2+}, respectively. As input, we used sequences spanning 135 nucleotides of the Hsp90 protein in yeast. To obtain these sequences, we added 54 nucleotides flanking both 5' and 3' sides of the region of interest (complete sequences were obtained from https://www.addgene.org/41188/sequences/). From each of these data sets, we selected the conformation with the lowest Gibbs free energy (dG) or with the highest melting temperature (Tm), as highest-stability reference points.

Since Hsp90 is a chaperone protein involved in the response to thermal stress as well as in the regulation of osmotic stress (Yang *et al.*, 2006; Boucher *et al.*, 2014), we tested the impact of each factor in each environment and amino-acid position and quantified how variation in temperature, osmotic stress and residue position affected the correlation between *mRNA melting temperature*, *codon frequency* and *Gibbs free energy* and the effect of synonymous mutations. We used the following models:

$$Y = \text{melting temperature} + \epsilon$$

$$Y = \text{codon frequency} + \epsilon$$

$$Y = \text{Gibbs free energy} + \epsilon$$

$$\begin{aligned} \text{Y} \quad &= \text{temperature} + \text{salt} + \text{residue position} + \text{codon frequency} + \text{Gibbs free energy} \\ &+ \text{melting temperature} + (\cdots) + \text{temperature*salt*residue depth*codon frequency} \\ &*\text{Gibbs free energy*melting temperature} + \epsilon \end{aligned}$$

where Y corresponds to the fitness effect of synonymous mutations (see above), *temperature* is a covariate coding for 25°C, 30°C and 36°C, *salt* is a fixed factor that codes for the presence or absence of added salt in the medium, *residue position* relates to the residue position in the amino-acid sequence (582 to 590), *codon frequency* is the frequency of each codon in the genome (see above), *Gibbs free energy* is the variation in Gibbs free energy obtained from mfold and *melting temperature* the melting temperature for the RNA estimated through mfold. We included all interactions between the studied factors in the model. The effect size of each term in the model was estimated through the η^2 (Levine and Hullett, 2002) for 1000 posterior samples.

All analyses were performed with R (R version 3.3.3) (R Core Team, 2017) or Mathematica 11 (version 11.2) (Wolfram Research, Inc., 2017). The complete documentation of all analyses, which allows for the reiteration of all steps, is available as Online Supplementary Material.

3 Results & Discussion

We implemented a software to infer selection coefficients from deep mutational scanning experiments. The *empiricIST* software is based on a previously developed Bayesian Markov chain Monte Carlo (MCMC) approach (Bank *et al.*, 2014), and is a user-friendly and accurate software for improved growth rate estimation from time-sampled deep-sequencing data. We took advantage of the high accuracy provided by this method to estimate selection coefficients for synonymous mutations.

3.1 Bayesian MCMC outperforms linear regression

Validating the method with various types of pseudo-data shows that our MCMC generally outperforms the ordinary least square regression (OLS). Figures 2 and S1 show the results for the standard and the data-based simulations (see Material & Methods). Although the mean square error (MSE) of the MCMC is comparable to that of the OLS when analyzing few time points (i.e., 3 to 5 time points), the MSE of the MCMC decreases faster as the number of time points increases (Fig. 2A).

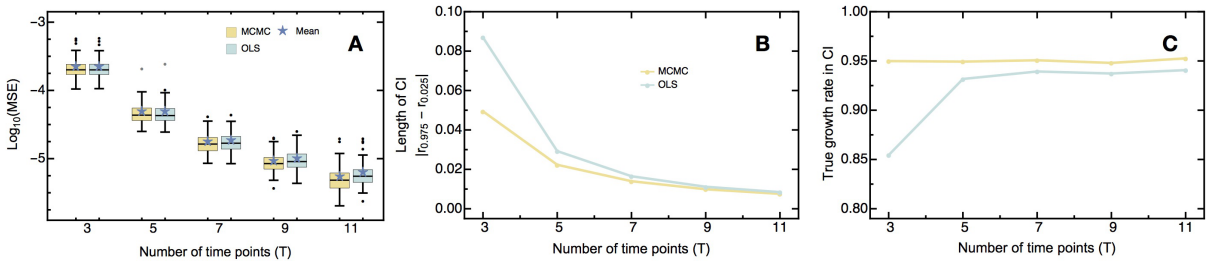

Figure 2: **Comparison between performance of *empiricIST* and ordinary least square regression with varying number of time points sampled.** A) mean square error (MSE), B) size of the credibility interval (CI), and C) the proportion the true growth rate contained in the CI. As shown, *empiricIST* shows equal or lower MSE than OLS regression, particularly as the number of sampled time points increases. Furthermore, *empiricIST* outperforms the OLS regression in terms of the size of the CI and in capturing the true growth rate, even when sampling a small number of time points.

Furthermore, when analyzing few time points, the length of the credibility interval (CI) is significantly smaller for the MCMC than the corresponding confidence interval of the OLS regression (Fig 2B). While the difference between the length of the confidence intervals decreases as the number of time samples T increases, the size of the CI from the MCMC always remains smaller, which implies that it delivers more precise and accurate results than the conventional OLS regression. Most importantly, and unlike the OLS regression, the CI of the MCMC remains well calibrated along the entire range of parameters (Fig. 2 C), despite being generally narrower than its OLS counterpart.

3.2 Synonymous mutations have detectable effects on fitness

Previous studies have shown that synonymous mutations can directly affect fitness (e.g. Lind *et al.*, 2010; Firnberg *et al.*, 2014; Hunt *et al.*, 2014) and impact the ability of populations to adapt to new environments (Bailey *et al.*, 2014; Agashe *et al.*, 2016). For example, Bailey *et al.* (2014) found that two synonymous mutations were driving adaptation to a new medium in two experimental replicates by increasing the expression of a gene involved in glucose metabolism.

In a more recent study, Agashe *et al.* (2016) found that the deleterious effect of synonymous mutations in a medium with methylamine as the sole carbon source could be rescued by different mutations, including four synonymous mutations that increased transcription and protein production levels. The impact of synonymous mutations at the genome wide level can also be found in patterns of codon usage bias (synonymous codons are used in different frequencies) across genomes. Evidence coming from studies within and between species support the role of direct selection on synonymous sites in various genes (DuMont *et al.*, 2004; Singh *et al.*, 2007; Hershberg and Petrov, 2009; Ran and Higgs, 2010; Shah and Gilchrist, 2011; Choi and Aquadro, 2016; Sun *et al.*, 2016). A first piece of evidence for synonymous effects in the studied region of Hsp90 came from Bank *et al.* (2014), who reported that one of the 15 mutations synonymous to the parental sequence had a significantly deleterious effect in 4 out of 6 environments (Fig. 9 Bank *et al.*, 2014). In order to evaluate the effects of synonymous mutations on a larger scale, we applied the *empiricIST* software to the data set from Bank *et al.* (2014), which consists of bulk competitions of all 576 possible single codon-changing mutations in a 9 amino-acid region of Hsp90 in *Saccharomyces cerevisiae* across 6 different experimental conditions. For the environments 30N and 30S (30°C with normal and high salinity) we confirmed our results across the available 3 and 2 replicates, respectively. To quantify the effect of synonymous mutations as compared with the effect of non-synonymous mutations and experimental error we estimated the absolute pairwise difference between random pairs of amino-acids, codons, posterior samples and replicates (Table S2, see Material & Methods). On average, the effects of synonyms are small, but larger than the experimental error. In fact, 11% of the synonymous mutations present in the 5% beneficial tail (normalized to the effect of each amino-acid, see Material & Methods) are common between the three 30N replicates and 11% between the two 30S replicates, as compared with 0.0125% vs. 0.025% overlap expected by chance. The median pairwise difference between two synonymous mutations was between 3% (in 25S) and 27% (in 36N) larger than the difference between two draws from the posterior of the same codon (Table S2). As expected, the average effect of synonymous mutations is much smaller than that of non-synonymous mutations (Table S2). The estimated average fitness difference between two synonymous mutations is between 13% (in 36N) to 32% (in 25S) of the difference between two non-synonymous mutations. In concordance with these estimates, a one-way ANOVA shows that 20-30% of the fitness changes within amino-acids can be explained by codon variation alone (see Fig. S8). On average, the effects of synonymous mutations are higher in the 36N environment (Table S2, Fig. S8), where Hsp90 is expected to be more important for organism survival (Yang *et al.*, 2006; Boucher *et al.*,

2014; Mishra *et al.*, 2016).

3.3 The beneficial tail of the distribution of synonymous fitness effects

The distribution of fitness effects contains information about the availability of beneficial mutations (Orr, 2005, 2010). It is of particular interest to study the shape of the beneficial tail of this distribution as it determines various aspects about the nature of adaptive walks (Orr, 2010; Eyre-Walker, 2006). Bank *et al.* (2014) previously found that for all of the environments, except for 25S, the beneficial tail of the full distribution of fitness effects most likely belonged to the Weibull domain. This suggested that populations were close to a well-defined optimum, and the available beneficial mutations would be of similar and small size (Orr, 2010; Joyce *et al.*, 2008; Bank *et al.*, 2014).

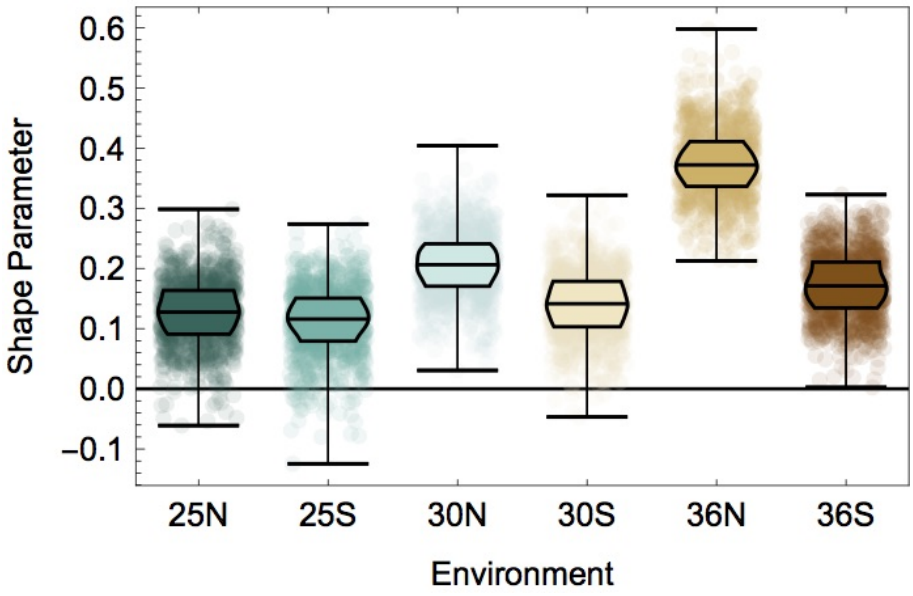

Figure 3: **Distribution of the shape parameter of the beneficial tail of synonymous mutations.** The shape parameter was estimated using the tail shape estimator from *empiricIST* and used the information of 1000 samples of the posterior distribution. In all the environments, the shape of the tail is clearly positive indicating that it belongs to the Fréchet domain. This implies that most mutations in this distribution will be characterized by nearly neutral effects. Environmental conditions are indicated by the combination of temperature (25C, 30C and 36C) and salinity (N = normal and S = high salinity).

We used the tail shape estimator from the *empiricIST* software to estimate the tail shape of the distribution of beneficial synonymous mutations. To obtain this distribution, we subtracted the average amino-acid effect from the selection coefficient of each codon. We find that the shape parameter of the fitted Generalized Pareto Distribution (GPD) is most likely positive in all environments, which indicates that the resulting shape of the beneficial tail is likely to

belong to the Fréchet domain (Fig. 3) (Orr, 2010; Joyce *et al.*, 2008). Distributions from this domain are characterized by many mutations of small effect, along with few mutations of large and unpredictable effect (Joyce *et al.*, 2008; Neidhart and Krug, 2011; Jain and Seetharaman, 2011). This is consistent with the expectation that most synonymous mutations in the whole data set have little effect on fitness, but some have large fitness effects.

3.4 Synonymous mutations affect the topography of the landscape

We investigated the effect of synonymous mutations on the topography of the fitness landscape by comparing averaged and single-effect landscapes (see Fig. 1 C, 1 D, Material & Methods) for each of the 9 amino-acid positions across 6 environments. For all 54 landscapes, we computed two statistics: the roughness-to-slope ratio r/s (Szendro *et al.*, 2013) and the locus-specific gamma statistic (Ferretti *et al.*, 2016). The roughness-to-slope ratio describes the prevalence of epistasis (i.e., the extent of non-linear fitness effects between mutations) in relation to the magnitude of fitness effects in the landscape (Carneiro and Hartl, 2010; Schenk *et al.*, 2013). The $\gamma_{i\to j}$ statistic measures the correlation of fitness effects of the same mutations in a single-step distance across all genetic backgrounds. Whereas the roughness-to-slope ratio describes the landscape by means of only two values, $\gamma_{i\to j}$ results in a detailed fingerprint of the fitness landscape that makes heterogeneity of epistasis in the landscape visible. In general all landscapes are highly epistatic ($r/s > 1$), with the magnitude of the roughness-to-slope ratio depending on amino-acid position and environment (Fig. S9). Single-effect landscapes are slightly more epistatic (higher ratio) than averaged landscapes, although this difference is in general small. In high salinity, the difference in the r/s ratio between amino-acid positions and between averaged and single-effect landscapes is larger (Fig. S9). The increased epistatic signal observed in these environments could be caused by the combination of low absolute growth rates observed in high salinity conditions (c.f. Table 1 in Bank *et al.*, 2014) and larger experimental uncertainty (S7) in this environment. This indicates that one should be cautious when interpreting the roughness-to-slope ratio across data sets, because it may be reflecting differences in growth rates and experimental error between environments, rather than genuine changes in the epistatic component of the landscape. Computing the $\gamma_{i\to j}$ statistic shows that averaged and single-effect landscapes tend to display the same type of epistasis within amino-acid position and environment (Fig. 4, Fig. S10). In general, magnitude and sign epistasis (when the effect and sign of a mutation depends on the genetic background where it appears) are prevalent and we observe

only few cases of reciprocal sign epistasis (both mutations switch sign when combined; Fig. S10). Since reciprocal sign epistasis is a necessary, but not a sufficient condition for multiple fitness peaks in a landscape (Poelwijk *et al.*, 2011), its low prevalence suggests that there should be few fitness peaks in both averaged and single-effect landscapes. In contrast to the results from the roughness-to-slope ratio, the $\gamma_{i\to j}$statistic shows smaller differences between environments. As this statistic computes results based on the correlation and not the effect size of fitness effects across genetic backgrounds, it is less sensitive to variation in growth rates and experimental error. In fact, most differences in the type of epistasis are found when comparing the order in which mutations occur. In particular, landscapes resulting from non-synonymous mutations ($\gamma_{1\to 2}$, $\gamma_{2\to 1}$) display in general strong epistasis (Fig. 4), compared to landscapes that include synonymous mutations ($\gamma_{1\to 3}$, $\gamma_{2\to 3}$, $\gamma_{3\to 1}$, $\gamma_{3\to 2}$). However, the presence of magnitude/sign epistasis in both $\gamma_{3\to 1}$, $\gamma_{3\to 2}$ landscapes suggests that synonymous mutations do not have the same effects across different amino-acids. Similar to what was observed in the roughness-to-slope ratio, there is variation across amino-acid positions for both non-synonymous and synonymous mutations.. Thus, the structure of the codon table (i.e., the existence of synonymous and non-synonymous mutations) imposes a strong general pattern of epistasis on the landscape. However, variation in this pattern across positions and environments indicates that every single amino-acid position has indeed a differently shaped fitness landscape.

3.5 Impact of synonymous mutations on adaptive walks

Including synonymous mutations changes the topography of the landscape, which may affect the accessibility of different mutational paths by creating additional peaks and sinks in the fitness landscape. To quantify the impact of synonymous effects on adaptive walks, we calculated the number of optima, the mean expected length of adaptive walks, and the variance in the number of steps for the single-effect and averaged landscapes. We based our calculation on the assumption of the strong-selection weak-mutation limit (Gillespie, 1984), in which evolution happens by means of sequential mutational changes that result in an adaptive walk that ends in a fitness peak (e.g. Orr, 2005; Schoustra *et al.*, 2009; Frank, 2014; Zagorski *et al.*, 2016). We define a fitness peak as any genotype with fitness higher than all single-step mutational neighbors. For averaged landscapes, in which all synonymous mutations are assigned equal fitness, we consider a fitness plateau spanned by synonymous codons as a single local optimum if all non-synonymous codons in a distance of a single nucleotide step have lower fitness (as in

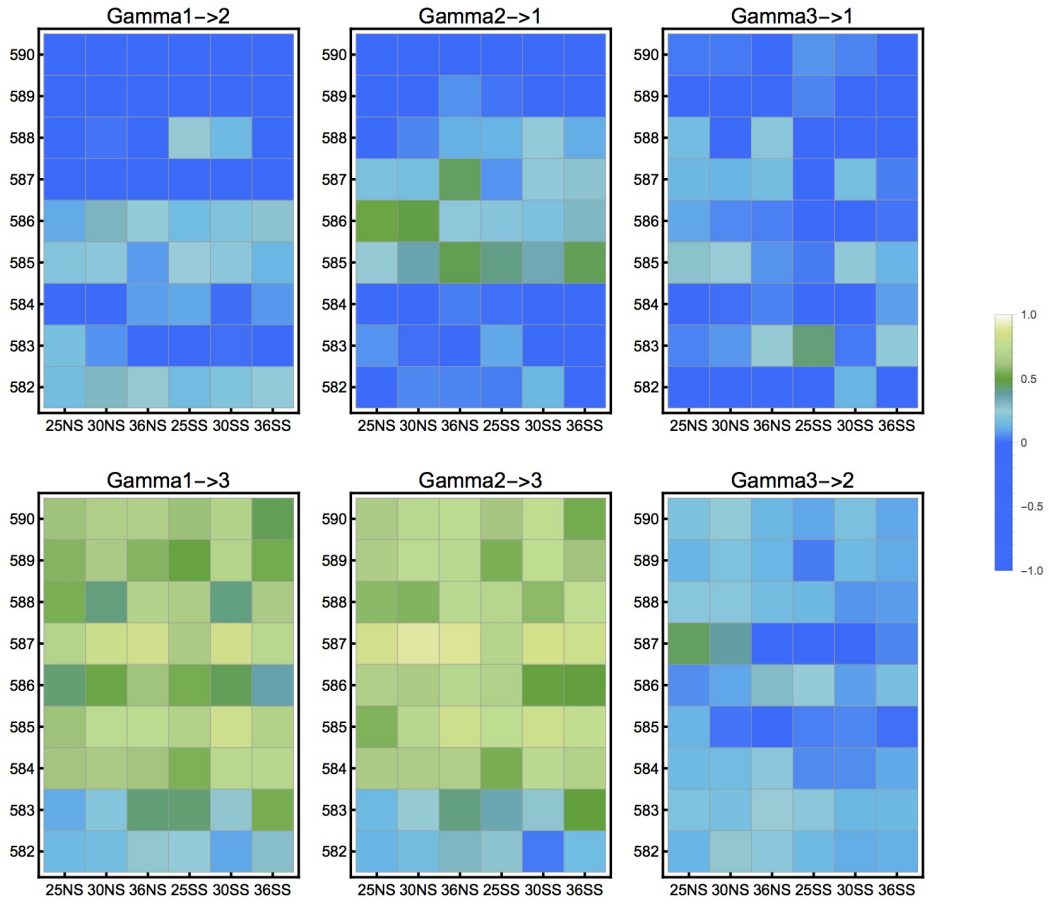

Figure 4: **Gamma statistic calculated for pairs of mutations in different codon positions for single-effect landscapes for the 9 amino-acid positions studied (y-axis) in the different environments (x-axis)**. In general, non-synonymous mutations on top of each other ($\gamma_{1\to2}$, $\gamma_{2\to1}$) show prevalence of sign epistasis (γ between 1 to -1/3), while non-synonymous mutations on top of synonymous mutations ($\gamma_{1\to3}$, $\gamma_{2\to3}$) show higher prevalence of magnitude epistasis (γ between 1 and 0). There is no clear variation across environments (x axis), but we find a clear impact of amino-acid position in the type of epistasis (y axis). Interestingly, both g$\gamma_{3\to1}$ and $\gamma_{3\to2}$ indicate a potential epistatic hotspot in positions 582 and 583 across all environments.

Fig. 1 C). By definition, the number of fitness peaks in the averaged landscape has to be lower or equal to that of the single-effect landscape. Indeed, we find that there is usually a large difference in the number of fitness peaks between single-effect and averaged landscapes (Fig. 5, Fig. S11). This difference is environment-dependent and also varies across amino-acid positions (Fig. 5, Fig. S11). The variation in the differences between single-effect and averaged landscapes is not consistent within buried or exposed positions (see Material & Methods), which suggests that the impact of synonymous mutations is not due solely to effects at the structural level of the protein. In contrast to what was expected due to the reduced number of observed peaks, we observe shorter and less variable adaptive walks for averaged landscapes than for single-effect

landscapes (Fig. S12, Fig. S13). This suggests that evolution on the 'true' landscape that includes effects of synonymous mutations is less predictable (Lobkovsky *et al.*, 2011; De Visser and Krug, 2014), and that it may stall at an intermediate optimum created by variation in synonymous fitness effects. Different environments leave a stronger signature in single-effect landscapes than in averaged landscapes. In fact, for most environments and positions, averaged landscapes show only 1 or 2 optima in the landscape (Fig. 5, Fig. S11). Conversely, among the single-effect landscapes the 25°C stands out with a large number of peaks, coupled with short adaptive walks (Fig. S12). This could reflect the lower constraint on Hsp90 function at low temperatures, as well as the lower absolute growth rates of the population under this condition which may open up more opportunities for adaptation. In further support for this hypothesis, we observe fewer optima and longer and more variable adaptive walks at 36N, which is in agreement with the importance of Hsp90 under high temperatures (Hietpas *et al.*, 2013; Bank *et al.*, 2014; Boucher *et al.*, 2014; Mishra *et al.*, 2016). This is consistent with the small number of beneficial mutations observed by Bank *et al.* (2014) under this condition. This points to a scenario in which there is increased evolutionary constraint, such that the number of solutions to the adaptive challenge is very limited.

Our results allow for an interesting thought experiment regarding the impact of synonymous mutations on evolution across populations of different sizes. The average small differences in synonymous effects observed here will be only be visible to selection in large populations, where they may frequently stall adaptation if the population gets trapped in a local fitness peak. Bottlenecks (i.e., sudden drops in the population size), which can occur under natural scenarios and are also frequently imposed in experiments, may render synonymous mutations effectively neutral and thus erase the difference between averaged and single-effect landscapes. By opening mutational paths and bridging fitness peaks, a (temporally) smaller population size could thus speed up adaptation and increase its predictability (Wright, 1931; Jain *et al.*, 2011). This effect, even if weak, would be in contrast to the slow down of adaptation and decrease of predictability of evolution in small populations proposed in standard population-genetic theory (Orr, 2000; Lanfear *et al.*, 2014).

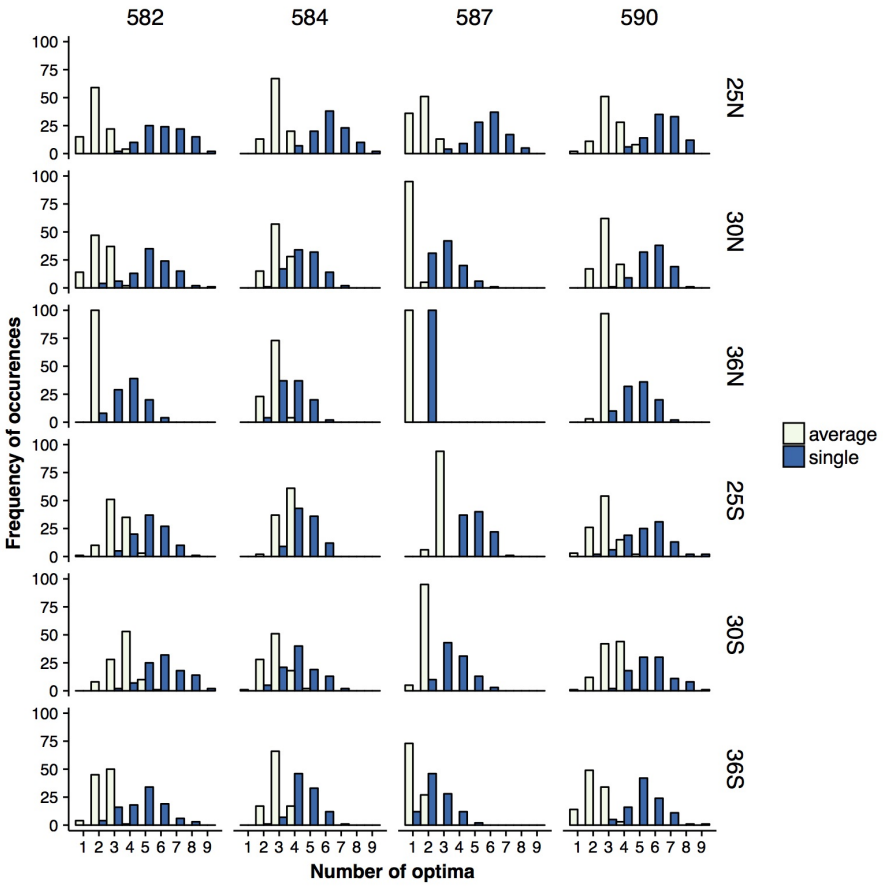

Figure 5: **Figure 5: Number of optima observed from 100 posterior samples of single-effect (dark blue) and averaged (light yellow) landscapes for positions 582, 584, 586 and 590 (from left to right) across environments.** The number of optima is always higher in single-effect than in averaged landscapes across environments. The number of optima is smaller at high temperatures, which may indicate increased constraints to adaptation. The large difference between the number of peaks in averaged and single landscapes suggests that synonymous mutations can affect adaptation to a new environment by trapping the population in a local optimum.

3.6 Effects of synonymous mutations are driven by a combination of mechanisms

Synonymous mutations can affect fitness by altering speed and accuracy of translation, and mRNA folding and stability (Drummond and Wilke, 2008; Kudla *et al.*, 2009; Zhou *et al.*, 2009; Sharp *et al.*, 2010; Plotkin and Kudla, 2011; Shabalina *et al.*, 2013; Presnyak *et al.*, 2015; Yu *et al.*, 2015; Knöppel *et al.*, 2016; Brule and Grayhack, 2017). It has been proposed that protein folding may be affected more significantly by changes in translation accuracy for buried (structural) positions, as they are often involved in the formation of crucial secondary and tertiary structures of the protein (Drummond and Wilke, 2008; Zhou *et al.*, 2009; Saunders and Deane, 2010). We evaluated whether the effect of synonymous mutations that we observe

can be explained by variation in codon preference or mRNA stability. For that, we analyzed a full linear model incorporating temperature, medium composition, residue position, Gibbs free energy, melting temperature of mRNAs, and codon usage frequency, as well as all possible interactions of those factors. We also estimated the slopes of linear regressions for each predictor to quantify the contribution of these factors to the fitness effects of synonymous mutations for each environment and amino-acid position. As may be expected, considering the diverse amino-acid positions and environments studied, no clear predictors of codon fitness emerged. We found fitness effects of synonymous mutations to be affected by interactions between residue positions and temperature, medium composition, mRNA melting temperature and Gibbs free energy, and codon usage frequency (Table S3). In addition, we saw a clear impact of residue position and temperature in the fitness effects of synonymous mutations (Table S3, Fig. S14). Despite generally weak correlations between fitness effects and the predictors (r^2 for codon frequency: 0.0235, r^2 for Gibbs free energy: 0.0217, r^2 for melting temperature: 0.0246), correlations tended to be stronger at higher temperatures in standard medium (Fig. S15). Namely, we observed that in 36N there is a deleterious effect of higher mRNA stability for positions 585 and 587, and a beneficial effect of higher mRNA stability for positions 583 and 584 (Fig. S15b,S15c). At this temperature, more common codons show beneficial effects at positions 586, 587 and 590, and negative effects at positions 583, 585 and 588. This indicates that the impact of mRNA stability and codon frequency on the fitness effects of synonymous mutations is dependent on residue position and environment. The usage of different synonymous codons allows cells to slow down or arrest protein production in response to sudden environmental changes and to optimize resource production (Zhang *et al.*, 2009; Fredrick and Ibba, 2010; Tuller *et al.*, 2010). Our results suggest that a combination of several mechanisms drives the effects of synonymous mutations. Namely, we found a correlation of fitness effects of synonymous mutations with mRNA stability and codon frequency. This is in line with other studies demonstrating that synonymous codons affect mRNA stability (Hilgers, 2006; Romero and Arnold, 2009; Presnyak *et al.*, 2015) by modulating protein translation kinetics through optimal codon usage (Akashi, 1994; Drummond and Wilke, 2008; Presnyak *et al.*, 2015; Harigaya and Parker, 2016). However, synonymous mutations do not show consistent effects within buried (586 to 590) or exposed positions (582 to 585). Finally, we observe that the effects of synonymous mutations and their predictors are environment-dependent, with stronger effects at high temperatures.

4 Conclusion

The impact of the codon table on the evolutionary dynamics on fitness landscapes has received little attention. This is a consequence of the vast size of the nucleotide space and the resulting fitness landscape dimensionality, which has led to most studies restricting themselves to the amino-acid level. This study demonstrates the importance of considering the codon level, because every single amino-acid position results in a fitness landscape that varies across environments. Using rate estimates obtained with *empiricIST*, a new software for the estimation of growth rates from deep mutational scanning data, we investigated the consequences of including synonymous mutations when characterizing the fitness landscape of single amino-acid positions across environments. Results demonstrate the extent to which synonymous mutations may impact the topography of the fitness landscape and affect adaptation, as well as their environmentally-dependent effects. The strongest effects are observed at high temperature, where the Hsp90 protein is likely under the strongest evolutionary constraint. Interestingly, we find support for a heavy-tailed distribution of beneficial synonymous effects across all environments, suggestive of mutations with many small effects, and few with potentially large effects. A structural analysis indicates that synonymous effects can be mediated by changes in mRNA stability and variation in codon preference. However, effects are strongly dependent on the residue position under study, which makes a clear identification of the predictors of synonymous effects difficult. Overall, this study demonstrates how synonymous mutations can directly impact both the path and endpoint of an adaptive walk, and thus highlights the importance of their consideration in the study of fitness landscapes.

5 Acknowledgments

We thank the members of the Bank lab for discussion of the manuscript. We thank Dan Bolon for our long-term collaboration as part of which the data were obtained originally. This work was supported by Fundação Calouste Gulbenkian and an ERC Starting Grant to JDJ.

References

Agashe D, Martinez-Gomez NC, Drummond DA, Marx CJ (2013). Good codons, bad transcript: Large reductions in gene expression and fitness arising from synonymous mutations in a key

enzyme. *Molecular Biology and Evolution* **30**: 549–560.

Agashe D, Sane M, Phalnikar K, Diwan GD, Habibullah A, Martinez-Gomez NC, *et al.* (2016). Large-Effect Beneficial Synonymous Mutations Mediate Rapid and Parallel Adaptation in a Bacterium. *Molecular Biology and Evolution* **33**: 1542–1553.

Aita T, Iwakura M, Husimi Y (2001). A cross-section of the fitness landscape of dihydrofolate reductase. *Protein Engineering, Design and Selection* **14**: 633–638.

Akashi H (1994). Synonymous codon usage in Drosophila melanogaster: Natural selection and translational accuracy. *Genetics* **136**: 927–935.

Bailey SF, Hinz A, Kassen R (2014). Adaptive synonymous mutations in an experimentally evolved Pseudomonas fluorescens population. *Nature Communications* **5**: 1–7.

Bali V, Bebok Z (2015). Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *International Journal of Biochemistry and Cell Biology* **64**: 58–74.

Bank C, Hietpas RT, Wong A, Bolon DN, Jensen JD (2014). A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: Uncovering the potential for adaptive walks in challenging environments. *Genetics* **196**: 841–852.

Bank C, Matuszewski S, Hietpas RT, Jensen JD (2016). On the (un)predictability of a large intragenic fitness landscape. *Proceedings of the National Academy of Sciences* **113**: 14085–14090.

Bataillon T, Zhang T, Kassen R (2011). Cost of adaptation and fitness effects of beneficial mutations in Pseudomonas fluorescens. *Genetics* **189**: 939–49.

Beisel CJ, Rokyta DR, Wichman HA, Joyce P (2007). Testing the extreme value domain of attraction for distributions of beneficial fitness effects. *Genetics* **176**: 2441–2449.

Boone EL, Merrick JR, Krachey MJ (2014). A Hellinger distance approach to MCMC diagnostics. *Journal of Statistical Computation and Simulation* **84**: 833–849.

Boucher JI, Cote P, Flynn J, Jiang L, Laban A, Mishra P, *et al.* (2014). Viewing protein fitness landscapes through a next-gen lens. *Genetics* **198**: 461–471.

Brule CE, Grayhack EJ (2017). Synonymous Codons: Choose Wisely for Expression. *Trends in Genetics* **33**: 283–297.

Carneiro M, Hartl DL (2010). Adaptive landscapes and protein evolution. *Proceedings of the National Academy of Sciences* **107**: 1747–1751.

Chevin LM (2011). On measuring selection in experimental evolution. *Biology letters* **7**: 210–3.

Choi JY, Aquadro CF (2016). Recent and Long-Term Selection Across Synonymous Sites in Drosophila ananassae. *Journal of Molecular Evolution* **83**: 50–60.

De Visser JAGM, Krug J (2014). Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**: 480–490.

Drummond DA, Wilke CO (2008). Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell* **134**: 341–352.

DuMont VB, Fay JC, Calabrese PP, Aquadro CF (2004). DNA variability and divergence at the Notch locus in Drosophila melanogaster and D. simulans: A case of accelerated synonymous site divergence. *Genetics* **167**: 171–185.

Eyre-Walker A (2006). The genomic rate of adaptive evolution. *Trends in Ecology and Evolution* **21**: 569–575.

Ferretti L, Schmiegelt B, Weinreich D, Yamauchi A, Kobayashi Y, Tajima F, *et al.* (2016). Measuring epistasis in fitness landscapes: The correlation of fitness effects of mutations. *Journal of Theoretical Biology* **396**: 132–143.

Firnberg E, Labonte JW, Gray JJ, Ostermeier M (2014). A comprehensive, high-resolution map of a Gene's fitness landscape. *Molecular Biology and Evolution* **31**: 1581–1592.

Fowler DM, Fields S (2014). Deep mutational scanning: A new style of protein science. *Nature Methods* **11**: 801–807.

Frank SA (2014). Generative models versus underlying symmetries to explain biological pattern. *Journal of Evolutionary Biology* **27**: 1172–1178.

Fredrick K, Ibba M (2010). How the sequence of a gene can tune its translation. *Cell* **141**: 227–229.

Gelman A, Roberts G, Gilks W (1996). Efficient metropolis jumping rules. In: *Bayesian statistics*, vol. 5, pp. 599–607.

Gillespie JH (1983). A simple stochastic gene substitution model. *Theoretical Population Biology* **23**: 202–215.

Gillespie JH (1984). Molecular Evolution Over the Mutational Landscape. *Evolution* **38**: 1116–1129.

Gorter FA, Aarts MGM, Zwaan BJ, de Visser JAGM (2018). Local Fitness Landscapes Predict Yeast Evolutionary Dynamics in Directionally Changing Environments. *Genetics* **1**: 307–322.

Harigaya Y, Parker R (2016). Analysis of the association between codon optimality and mRNA stability in Schizosaccharomyces pombe. *BMC Genomics* **17**: 1–16.

Hershberg R, Petrov DA (2009). General rules for optimal codon choice. *PLoS Genetics* **5**: 1–10.

Hietpas R, Roscoe B, Jiang L, Bolon DN (2012). Fitness analyses of all possible point mutations for regions of genes in yeast. *Nature Protocols* **7**: 1382–1396.

Hietpas RT, Bank C, Jensen JD, Bolon DNA (2013). Shifting fitness landscapes in response to altered environments. *Evolution* **67**: 3512–3522.

Hietpas RT, Jensen JD, Bolon DNA (2011). Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences* **108**: 7896–7901.

Hilgers V (2006). Translation-independent inhibition of mRNA deadenylation during stress in Saccharomyces cerevisiae. *RNA* **12**: 1835–1845.

Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C (2014). Exposing synonymous mutations. *Trends in Genetics* **30**: 308–321.

Jain K, Krug J, Park SC (2011). Evolutionary advantage of small populations on complex fitness landscapes. *Evolution* **65**: 1945–1955.

Jain K, Seetharaman S (2011). Multiple adaptive substitutions during evolution in novel environments. *Genetics* **189**: 1029–1043.

Joyce P, Rokyta DR, Beisel CJ, Orr HA (2008). A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation. *Genetics* **180**: 1627–1643.

Knöppel A, Näsvall J, Andersson DI (2016). Compensating the Fitness Costs of Synonymous Mutations. *Molecular Biology and Evolution* **33**: 1461–1477.

Kudla G, Murray AW, Tollervey D, Plotkin JB (2009). Coding-Sequence Determinants of Gene Expression in Escherichia coli. *Science* **324**: 255–258.

Lanfear R, Kokko H, Eyre-Walker A (2014). Population size and the rate of evolution. *Trends in Ecology and Evolution* **29**: 33–41.

Levine TR, Hullett CR (2002). Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research* **28**: 612–625.

Lind PA, Berg OG, Andersson DI (2010). Mutational robustness of ribosomal protein genes. *Science* **330**: 825–827.

Lobkovsky AE, Wolf YI, Koonin EV (2011). Predictability of evolutionary trajectories in fitness landscapes. *PLoS Computational Biology* **7**: 1–11.

Lüdecke D (2017). *sjstats: Statistical Functions for Regression Models*. R package version 0.14.0. **URL:** *https://CRAN.R-project.org/package=sjstats*

Markham NR, Zuker M (2008). UNAFold: Software for nucleic acid folding and hybridization. *Methods in Molecular Biology* **453**: 3–31.

Matuszewski S, Hildebrandt ME, Ghenu AH, Jensen JD, Bank C (2016). A statistical guide to the design of deep mutational scanning experiments. *Genetics* **204**: 77–87.

Mishra P, Flynn JM, Starr TN, Bolon DN (2016). Systematic Mutant Analyses Elucidate General and Client-Specific Aspects of Hsp90 Function. *Cell Reports* **15**: 588–598.

Neidhart J, Krug J (2011). Adaptive walks and extreme value theory. *Physical Review Letters* **107**: 1–4.

O'Brien EP, Ciryam P, Vendruscolo M, Dobson CM (2014). Understanding the influence of codon translation rates on cotranslational protein folding. *Accounts of Chemical Research* **47**: 1536–1544.

Orr HA (2000). The rate of adaptation in asexuals. *Genetics* **155**: 961–968.

Orr HA (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics* **6**: 119–127.

Orr HA (2010). The population genetics of beneficial mutations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1195–1201.

Parmley JL, Hurst LD (2007). How do synonymous mutations affect fitness? *BioEssays* **29**: 515–519.

Perfeito L, Ghozzi S, Berg J, Schnetz K, Lässig M (2011). Nonlinear Fitness Landscape of a Molecular Pathway. *PLoS Genetics* **7**: e1002160.

Plotkin JB, Kudla G (2011). Synonymous but not the same: The causes and consequences of codon bias. *Nature Reviews Genetics* **12**: 32–42.

Poelwijk FJ, Tănase-Nicola S, Kiviet DJ, Tans SJ (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *Journal of Theoretical Biology* **272**: 141–144.

Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, *et al.* (2015). Codon optimality is a major determinant of mRNA stability. *Cell* **160**: 1111–1124.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. **URL:** *https://www.R-project.org/*

Ran W, Higgs PG (2010). The influence of anticodon-codon interactions and modified bases on codon usage bias in bacteria. *Molecular Biology and Evolution* **27**: 2129–2140.

Rodriguez-Verdugo A, Carrillo-Cisneros D, Gonzalez-Gonzalez A, Gaut BS, Bennett AF (2014). Different tradeoffs result from alternate genetic adaptations to a common environment. *Proceedings of the National Academy of Sciences* **111**: 12121–12126.

Rokyta DR, Beisel CJ, Joyce P, Ferris MT, Burch CL, Wichman HA (2008). Beneficial Fitness Effects Are Not Exponential for Two Viruses. *Journal of molecular evolution* **67**: 368–376.

Romero PA, Arnold FH (2009). Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10**: 866–876.

Sauna ZE, Kimchi-Sarfaty C (2011). Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics* **12**: 683–691.

Saunders R, Deane CM (2010). Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Research* **38**: 6719–6728.

Schenk MF, Szendro IG, Salverda MLM, Krug J, De Visser JAGM (2013). Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. *Molecular Biology and Evolution* **30**: 1779–1787.

Schoustra SE, Bataillon T, Gifford DR, Kassen R (2009). The properties of adaptive walks in evolving populations of fungus. *PLoS Biology* **7**: 1–10.

Shabalina SA, Spiridonov NA, Kashina A (2013). Sounds of silence: Synonymous nucleotides as a key to biological regulation and complexity. *Nucleic Acids Research* **41**: 2073–2094.

Shah P, Gilchrist MA (2011). Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proceedings of the National Academy of Sciences* **108**: 10231–10236.

Sharp PM, Emery LR, Zeng K (2010). Forces that influence the evolution of codon bias. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1203–1212.

Singh ND, Bauer DuMont VL, Hubisz MJ, Nielsen R, Aquadro CF (2007). Patterns of mutation and selection at synonymous sites in Drosophila. *Molecular Biology and Evolution* **24**: 2687–2697.

Sun Y, Tamarit D, Andersson SG (2016). Switches in Genomic GC Content Drive Shifts of Optimal Codons under Sustained Selection on Synonymous Sites. *Genome Biology and Evolution* p. evw201.

Szendro IG, Schenk MF, Franke J, Krug J, De Visser JAGM (2013). Quantitative analyses of empirical fitness landscapes. *Journal of Statistical Mechanics Theory and Experiment* p. P01005.

Tuller T, Waldman YY, Kupiec M, Ruppin E (2010). Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences* **107**: 3645–3650.

Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**: 111–114.

Wenger JW, Piotrowski J, Nagarajan S, Chiotti K, Sherlock G, Rosenzweig F (2011). Hunger artists: Yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLoS Genetics* **7**: 1–17.

Wolfram Research, Inc (2017). Mathematica v. 11.2. Champaign, Illinois, USA.

Wright S (1931). Evolution in mendelian populations. *Genetics* **16**: 97–159.

Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R (2016). Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**: 1–21.

760 Yang XX, Maurer KCT, Molanus M, Mager WH, Siderius M, Van Der Vies SM (2006). The
761 molecular chaperone Hsp90 is required for high osmotic stress response in Saccharomyces
762 cerevisiae. *FEMS Yeast Research* **6**: 195–204.

763 Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, *et al.* (2015). Codon Usage Influences the
764 Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular*
765 *Cell* **59**: 744–754.

766 Zagorski M, Burda Z, Waclaw B (2016). Beyond the Hypercube: Evolutionary Accessibility of
767 Fitness Landscapes with Realistic Mutational Networks. *PLoS Computational Biology* **12**:
768 1–18.

769 Zhang G, Hubalewska M, Ignatova Z (2009). Transient ribosomal attenuation coordinates protein
770 synthesis and co-translational folding. *Nature Structural and Molecular Biology* **16**: 274–280.

771 Zhou T, Weems M, Wilke CO (2009). Translationally optimal codons associate with structurally
772 sensitive sites in proteins. *Molecular Biology and Evolution* **26**: 1571–1580.

773 Zuker M, Mathews D, Turner D (1999). Algorithms and Thermodynamics for RNA secondary
774 structure prediction: a pratical guide. *RNA biochemistry and biotechnology* pp. 1–33.