The distribution of fitness effects of synonymous mutations in a gene under selection

Eleonore Elisabeth Lebeuf-Taylor

Thesis submitted to
The Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements for the
M.Sc. degree in the
Department of Biology

University of Ottawa

ABSTRACT

Models of adaptive evolution consider the path that a population takes to a fitness optimum through a series of beneficial single-nucleotide changes, although synonymous mutations, which do not alter the amino acid sequence, have been conventionally seen as 'silent' changes. However, recent studies have reported the existence of spontaneous adaptive synonymous substitutions, raising the possibility that this type of mutation is an overlooked source of adaptive potential. Using site-directed mutagenesis, we constructed and quantified the Darwinian fitness of 40 synonymous and 71 nonsynonymous individual mutants of a gene essential for glucose uptake in *Pseudomonas fluorescens*. The gene in question, *gtsB*, is the target of natural selection in an environment where glucose is the growth-limiting factor. Of the 56 beneficial mutations, half were synonymous and covered as wide a range of beneficial fitness effects as the nonsynonymous sample. Drawing on extreme value theory, we show that the fitness effects of synonymous mutations correspond to a distribution with a right-truncated tail belonging to the Weibull domain of the Generalised Pareto Distribution. Moreover, the distribution of beneficial fitness effects of synonymous mutations is indistinguishable from that of nonsynonymous mutations, revealing their previously unrecognised adaptive potential.

RÉSUMÉ

Selon les modèles d'évolution adaptative, la trajectoire qu'emprunte une population pour arriver à une valeur sélective optimale passe par une série de mutations ponctuelles bénéfiques. Les mutations synonymes, qui n'entraînent pas de changement d'acide aminé, ont longtemps été considérées comme étant silencieuses. Toutefois, la découverte récente de mutations synonymes bénéfiques ayant émergé spontanément suggère que ce type de mutation possède un potentiel adaptatif sous-estimé. Grâce à la mutagénèse dirigée, nous avons créé et quantifié la valeur sélective de 40 lignées synonymes et 71 lignées non-synonymes chez *Pseudomonas fluorescens*, les mutations touchant un gène essentiel à la capture du glucose. Il s'agit du gène *gtsB*, qui est visé par la sélection naturelle dans un milieu où la croissance de l'organisme n'est limitée que par la disponibilité du glucose. La moitié des 56 mutations bénéfiques étaient synonymes et représentaient une gamme de valeurs sélectives bénéfiques aussi large que celle de l'échantillon non-synonyme. Guidés par la théorie des valeurs extrêmes, nous démontrons que les effets sur la valeur sélective des mutations synonymes correspondent à une distribution ayant une queue bornée appartenant au domaine Weibull de la distribution généralisée de Pareto. Notamment, les distributions des effets sur la valeur sélective des mutations synonymes et non-synonymes bénéfiques sont indiscernables l'une de l'autre, suggérant que les mutations synonymes possèdent un potentiel adaptatif insoupçonné.

ACKNOWLEDGEMENTS

This thesis is the culmination of a collective endeavour involving many hands, and its success the fruit of extraordinary collaborations. First and foremost, I'd like to express my deep gratitude toward my supervisor, Dr. Rees Kassen, whose contagious passion for evolutionary biology and strong mentorship inspire me and have molded me into a budding, and lifelong, scientist, no matter where the meanders of life will take me. His support, on both scientific and personal levels, helped me face adversity with confidence and determination.

To Nick McCloskey, who kept this work going while I faced down cancer, I am incredibly grateful. I am also deeply indebted to Dr. Aaron Hinz, for conceiving the experimental framework that made this work possible, and Dr. Susan Bailey, for putting complex statistical investigations within my reach. To my fellow lab- and officemates, Alanna Leale, Jonathan Dench, Patrick Chen, Felipe Dargent, Sonal Shewaramani, Alana Schick, Chris Angell, thank you for your invaluable advice, insight, companionship, and our entertaining and enlightening tangents on exotic animals, Chinese zodiac and extraordinary *Drosophila* sperm. I would also like to extend thanks to my advisory committee members, Dr. Risa Sargent, Dr. Vincent Careau and Dr. Nicolas Rodrigue, for their sound advice and encouragement throughout this journey of scientific learning and discovery.

On a personal note, I would like to thank my friends around the world, in particular Ilona Clocher and Gabrielle Leduc-Lebeuf, on whose friendship and love I could always count to elevate my spirits during the hard times and to share in a multitude of joys. Thank you to my partner, Felix Glaser, a grounding force whose grit and integrity continually inspire me, and with whom adventuring in the wild is always on the horizon.

I could not have succeeded in both defeating disease and returning to complete my Master's without the unconditional and unwavering love, support, and care of my parents, Glen and Christine, and my sister, Isabelle. Du plus profond de mon cœur, merci. Your patience and unfaltering enthusiasm as I contemplate a thousand and one life projects are extraordinary.

My research was funded by an NSERC-CGSM scholarship, in combination with the Department of Biology - University of Ottawa.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THESIS STATEMENT

Given the health challenges that I started to face shortly after starting my Master's, this work is the fruit of many collaborations. Part of the experimental work for the construction of a point mutant library was performed by Nick McCloskey. Throughout my thesis, the use of plural first person reflects the many hands that brought this project to fruition, with help from Dr. Julian Evans in producing Figure 4.

CHAPTER 1: GENERAL INTRODUCTION

I. The sounds of 'silent' mutations

Evolution occurs when there is a change in allelic frequency over time, which is dependent on the existence of genetic variability. The point mutation, which modifies a single letter in a DNA sequence, represents the most basic unit of genetic variability and although minute, this deviation from a given sequence can have important effects in terms of adaptation or disease (Chamary et al. 2006; Cuevas et al. 2012; Hunt et al. 2014). Owing to the degeneracy of the genetic code, all amino acids – apart from methionine and tryptophan – can be encoded by two to six codons. This leads to two classes of point mutations: those that change the peptide sequence, known as nonsynonymous mutations, and those that do not, known as synonymous mutations. Conventionally, only nonsynonymous mutations have been considered in studies of mutational effects; their effects are generally deleterious. On the other hand, synonymous mutations, which do not alter the encoded amino acid, have been conventionally referred to as 'silent' mutations. The range and relative frequency of their phenotypic effects, unlike those of nonsynonymous mutations, have not yet been characterised.

Natural selection acts on this variability in phenotypes to drive evolution. To model evolution, however, we must bridge the gap between changes to a genetic sequence and differences in phenotype. This can be done by considering the distribution of fitness effects (DFE) of available mutations. A DFE is based on the relative frequencies and effect sizes of deleterious, neutral and beneficial mutations. Microbial systems, which can be genetically manipulated easily and grown to several generations in a short period of time, offer an ideal system to test the fitness effects of point mutations. The DFE of nonsynonymous mutations,

which has been characterised – albeit in the genomes of microorganisms and in coding regions of eukaryotic genomes – is multimodal, with a primary peak around a weakly deleterious mode and a secondary peak of strongly deleterious effects (Eyre-Walker and Keightley 2007; Hietpas et al. 2011; Bank et al. 2014; Firnberg et al. 2014; Li et al. 2016). These observations are consistent with the prediction that the wild-type (WT) is well adapted to its environment and that most mutations therefore reduce fitness and are purged by selection. However, only beneficial mutations, which are rare, can be adaptive and thus constitute the raw material of natural selection. Consequently, developing a predictive model of adaptive evolution requires understanding the underlying DFE from which point mutations are drawn – specifically, to characterise both the deleterious and beneficial sides, subject to purifying and positive selection respectively.

Two questions are central to our study: What is the frequency of adaptive mutations of a given effect size? More specifically, what is the DFE of adaptive synonymous mutations? Attempts to answer the first question have been framed in terms of nonsynonymous mutations: in a comprehensive review of the DFE of new mutations, no mention at all is made of the possible DFE of synonymous mutations (Eyre-Walker and Keightley 2007). Nevertheless, there is clear evidence they are far from universally neutral: synonymous mutations have been shown to face purifying selection in diverse systems, from an RNA virus (an incidental observation from a large bank of single-nucleotide substitutions) (Sanjuán et al. 2004), to the metazoan *Drosophila melanogaster* (a genome-wide analysis showed constraints on synonymous sites) (Lawrie et al. 2013), to the bacterium *Salmonella typhimurium* (synonymous mutations in two non-essential ribosomal proteins had an average growth rate of 0.92 relative to the ancestor) (Lind et al. 2010).

Additionally, work in the proteobacterium *Methylobacterium extorquens* demonstrated that experimentally synthesised synonymous variants of an essential metabolic enzyme had reduced fitness compared with the WT (Agashe et al. 2013). This study targeted the *fae* gene, which encodes a protein in the formaldehyde oxidation pathway, essential for growth on methanol-only media. Its substrate specificity and high expression level render it useful for fitness measurements. Although this study did not measure the fitness contribution of individual mutations, with each allele containing between 46 and 150 single nucleotide substitutions, the reduced growth rate of all mutant strains compared with the WT clearly indicates that synonymous mutations in prokaryotes can also be under negative selection.

The evidence that synonymous mutations can face positive selection is sparser. In follow-up work to the *M. extorquens* study, the synonymous strains were allowed to evolve and made fitness gains – approaching WT levels – through compensatory mutations in either the *fae* coding sequence or its directly upstream intergenic region (Agashe et al. 2016). Out of nine coding region single nucleotide polymorphisms (SNPs), four were synonymous and evolved in parallel in eight replicate populations. The large beneficial fitness effects and high level of repeatability of these synonymous SNPs suggest that beneficial synonymous mutations likely play an important and underappreciated role in the adaptive process. However, it is worth noting that these beneficial mutations arose in the background of fitness-reducing variation, and their beneficial effects may be due to epistatic interactions with the previously introduced mutations. They are nonetheless representative of how compensatory mutations can be positively selected when in a reduced-fitness background. A synonymous mutation has also been shown to provide antifungal resistance in yeast (She and Jarosz 2018), providing yet more evidence that synonymous mutations may be underappreciated contributors to adaptation.

The impetus for this project comes from a study by Bailey et al. (2014), who reported the discovery of two spontaneous, highly beneficial synonymous mutations arising independently over the course of a selection experiment in *Pseudomonas fluorescens* SBW25. The mutations increased relative fitness by 7.25% and 8.73%, on par with the effect of a third independently arising nonsynonymous mutation (Bailey et al. 2014). All three mutations were located in the *gtsB* gene, which encodes a protein essential to glucose uptake, and went to near-fixation within about 200 generations, the sign of strong positive selection. Initially isogenic, the SBW25 population was perpetuated for over 1000 generations with glucose as the only carbon source in minimal media; mutations to *gtsB* show that it is the target of selection in an environment where glucose is a growth-limiting factor.

These observations are striking in that positively selected synonymous substitutions might be much more common than previously thought. In this experiment, all three mutations clustered near the start of the gene: at codons 10 (alanine → threonine), 15 (alanine; CGA → CGC), and 38 (glycine; GGC → GGT). Further genomic analysis showed that in closely related Pseudomonads, the observed synonymous mutation CGA → CGC at codon 15 was the most common substitution, occurring in 17/23 related strains (representing, parsimoniously, 3 independent changes). However, only 1/23 related strains showed a substitution at codon 38, although this was the same synonymous mutation (GGC → GGT) that occurred in *SBW25*.

The *gtsB* gene encodes a permease subunit of an ATP-binding cassette that is crucial to the transport of glucose through the inner membrane into the cell. It is 909 base pairs long and encodes 302 amino acids. When the synonymous mutant strains were grown on other carbon sources (mannitol and succinate), *gtsB(A15A)* and *gtsB(G38G)* had no effect on fitness; a knock-out of *gtsB* was also assayed and shown to decrease fitness dramatically to around 0.85 relative

to the ancestor, but had no effect when grown on other sugars. These tests demonstrate that *gtsB* is the target of selection on media containing glucose as the only carbon source and growth-limiting factor. Taken together, these observations provided a biologically relevant experimental framework under which we would seek answers to the following questions: first, what is the shape of the DFE of synonymous mutations? Secondly, how common are synonymous mutations with highly beneficial effects, such as those observed in the Bailey study? In other words, are synonymous mutations generally neutral, or is their DFE similar to that of nonsynonymous mutations? Showing the latter to be true would turn the assumption that these mutations are silent on its head.

II. Measuring fitness

In general terms, fitness in the traditional sense can be defined as an organism's genetic contribution to future generations. Fitness is difficult to measure in non-microbial systems, as it represents the combination of multiple phenotypic traits, and consequently a single mutation's effect on fitness is near impossible to detect. Moreover, as fitness is a measure of an organism's lifetime reproductive success, researchers commonly rely on proxies, such as number of offspring or seeds produced, to guide fitness estimates. Fortunately, many of these limitations can be overcome through methods of experimental evolution that exploit microbes to study evolutionary processes. Microbes have short generation times, can be grown as multiple replicate populations in manipulable environment, and can be routinely frozen to allow direct comparison of ancestral and derived populations, effectively allowing us to track the evolutionary trajectory of a given genotype.

Most importantly, experimental evolution gives us the tools to measure fitness as a phenotype. This is done by co-culturing two strains and measuring their relative contribution to future generations after a short timespan – an approach known as a competition – yielding an estimate of relative fitness. Rather than absolute fitness, this estimate can be considered "gene fitness", since it measures the ability of a gene to accomplish a given function, with a corresponding effect on growth, and holistically encompasses mutational effects at the DNA, RNA and protein level (Firnberg et al. 2014).

These competition assays are a common approach to measuring fitness in experimental evolution and have notably been employed by Richard Lenski to compare genotypes of independently evolved *Escherichia coli* populations in his Long-Term Evolution Experiment (Wiser and Lenski 2015), as well as many others (Sanjuán et al. 2004; Cuevas et al. 2012; Lind and Andersson 2013; Hauber et al. 2016; Knöppel et al. 2016). Importantly, competitions wholly encompass the properties of bacterial culture such as lag time and exponential growth (Wiser and Lenski 2015). It has been shown that single mutations can accurately be measured when their fitness effect is greater than 1%, or in terms of relative fitness, if their relative fitness is greater than 1.01 (Eyre-Walker and Keightley 2007).

III. Mutational models of adaptive evolution

Although evolution by natural selection is the only mechanism that can produce adaptation, we still lack a thorough, quantitative framework to predict the genetics of adaptive evolution. Developing a genetic theory of adaptation involves understanding how often, and to what extent, individual mutations are beneficial. Attempts to describe the shape of the underlying distribution from which point mutations are drawn began in the 1930s with Fisher, who proposed a

theoretical model in which individual mutations had different effect sizes on a trait, in what is now referred to as Fisher's geometric model (FGM) (Fisher 1930). FGM provides a framework for understanding how mutations that affect a phenotype impact fitness, and predicts that adaptation occurs through beneficial mutations with very small effect sizes, those with larger effect sizes being highly improbable. Reconciling Fisher's predictions with the genetic basis of adaptation – that is, considering changes as mutations in DNA rather than traits, Maynard Smith proposed that a given sequence occupies a discrete phenotypic space. For a gene of length L base pairs, assuming low mutation rates, there are only $3L$ possible point mutants (Maynard Smith 1970). Further, natural selection will only favour changes that increase the organism's fitness, bringing it closer to an optimum.

These step-wise genetic changes can continue as long as a fitter genotype exists within the available space until the population reaches an optimum. But what is the underlying distribution of fitness values to which these mutations belong? Here, Gillespie's mutational landscape model (MLM) comes in. Gillespie argued that, since adaptation only "sees" beneficial mutations, the distribution of neutral and deleterious mutations is irrelevant to models of adaptive evolution (Gillespie 1983, 1984). Under the MLM, the WT is assumed to already enjoy high fitness; thus, Gillespie proposed that genetic models of adaptation should consider only the extreme right tail of DFEs. His insight was to draw on extreme value theory (EVT), a branch of statistics that deals with the values located in the tails of distributions, to describe the distribution of beneficial mutations.

Further theoretical work has focused on characterising the behaviour of this distribution. As the magnitude of a mutation's effect increases, its frequency of occurrence is predicted to decrease, although the nature of this relationship is unclear. The distribution of extreme values

above a threshold – in this case, the WT fitness – assumes a form given by the Generalised Pareto Distribution (GPD), which encompasses three parent distributions. These are defined by the shape parameter ξ, with $\xi > 0$ corresponding to the Fréchet, $\xi = 0$ to the Gumbel, and $\xi < 0$ to the Weibull parent distributions; the tails are respectively heavy, light and bounded. Although the Gumbel family of distributions, which includes exponential and logistic distributions, has long been regarded as 'the' DFE of adaptive mutations, emerging experimental work is challenging this view. In fact, in Fisher's model, a population cannot adapt further once it has reached an optimum, lending weight to a predicted truncated tail of beneficial effects defined by the Weibull domain (Martin and Lenormand 2008).

Accordingly, while some theoretical models favour an exponential distribution (which falls under the Gumbel parent distribution), others propose tails with different weights (Orr 2005; Martin and Lenormand 2008). Similarly, there is empirical support for both the exponential distribution (Imhof and Schlotterer 2001), albeit as an imperfect fit (Kassen and Bataillon, 2006), and the Weibull distribution (Rokyta et al. 2008; Betancourt 2009; Schoustra et al. 2009; Bataillon et al. 2011), which puts an upper bound on fitness effects. Although microbial experimental evolution provides an ideal framework to directly observe the effects of individual mutations, the very rarity of beneficial mutations, coupled with their vulnerability to loss through drift and clonal interference, has made empirical testing of these models difficult. Moreover, small-effect mutations are hard to detect in selection experiments, where only mutations that confer a large fitness effect become fixed and therefore detectable (Rozen et al. 2002). Finally, as Martin et al. demonstrated, picking random mutants from a selection experiment can also provide an inaccurate representation of the DFE of point mutations as mutational hotspots may engender pseudoreplication (in this study, only a quarter of sequenced mutants were unique)

(McDonald et al. 2011). As we will detail further on, constructing clones of a fitness-relevant phenotype in a specific environment is a way to circumvent the need to wait for beneficial mutations to arise and fix in an evolving population, which inevitably yields small sample sizes (Rokyta et al. 2005).

IV. Phenotypic effects of synonymous mutations

i. Codon usage bias

Non-random use of synonymous codons, a phenomenon known as codon usage bias (CUB), has been widely noted across a wide range of taxa (Ikemura 1985; Sharp et al. 1988; Hershberg and Petrov 2008). Codon usage modulates translation rate (Sorensen et al. 1989), as preferred codons are thought to result in more efficient and accurate translation (Gingold and Pilpel 2011; Yu et al. 2015). Furthermore, preferred codons are overrepresented in highly expressed genes (Hilterbrand et al. 2012) and often, but not always, correspond to the relative availability of cognate tRNAs (Gouy and Gautier 1982; Duret 2000; Kanaya et al. 2001). Consequently, rare codons are generally thought to be translated more slowly due to the lower levels of corresponding tRNA and weaker mRNA-tRNA interaction at the third nucleotide position, known as the wobble position (Chaney and Clark 2015). A measure of codon preference is the Codon Adaptation Index (CAI), based on the relative abundance of synonymous codons in highly expressed genes, which are assumed to be under strong selective pressure for maximal translational efficiency and accuracy (Sharp and Li 1987). The tRNA Adaptation Index (dos Reis et al. 2004) further takes into account the availability of tRNAs in the cell and the strength of the codon-anticodon interaction, as the wobble position can lead to non-cognate mRNA-tRNA pairing.

CUB plays an important role in modulating mRNA stability (Presnyak et al. 2015) and, through its effect on translational dynamics, co-translational protein folding (Yu et al. 2015). Moreover, there appears to be selection for less-preferred codons at the start of genes; a slow start to elongation is thought to increase transcriptional efficiency (Plotkin and Kudla 2011). Experimental work in the fungus *Neurospora* showed a strong association between codon optimisation and both mRNA and protein expression levels (the two had a comparable fold increase) (Zhou et al. 2016). On the other hand, studies in mutated ribosomal protein genes in *S. typhimurium* failed to show any correlation between relative codon usage and the fitness effects of synonymous mutations (Lind et al. 2010). Notably, both beneficial synonymous mutations observed in the *P. fluorescens* study resulted in a change to slightly less-preferred codons, which suggests that alignment with CUB is not a primary mechanism for the beneficial fitness effect of these mutations (Bailey et al. 2014).

Observations from the *gtsB* study are reflected in recent work in *E. coli* that tested whether the same synonymous mutation at different codon positions within a ribosomal subunit gene (*rplQ*) has the same effect, as would be predicted if CUB were the only mechanism by which synonymous mutations have a fitness effect (Hauber et al. 2016). At five sites within the *rplQ* gene, the wild-type codon was changed to a less-preferred codon and the resultant synonymous single-mutants were competed against the wild-type. Four sites were paired: two lysine and two leucine sites were changed from a wild-type to a less-preferred codon. Two of the synonymous single-mutants exhibited fitness values of 1.004 and 1.006, which are very slightly higher relative to the ancestor, whereas the other three single-mutants had fitness values around 0.97, slightly lower than wild-type. Notably, fitness effects (beneficial or deleterious) were different at the two different sites for the matching lysine and leucine codons. Moreover, a mutant strain

containing all five synonymous mutations exhibited dramatically reduced fitness – approximately 20% lower than the WT.

Changes in codon preference can have large phenotypic effects. In one notable instance, a synonymous mutation to a rare codon changed the protein's substrate specificity, which was thought to be caused by altered co-translational folding (Kimchi-Sarfaty et al. 2007). Yet another study that tested synonymous variants of green fluorescent protein (GFP) found no correlation between CAI and fluorescence levels (Kudla et al. 2009). While these observations contribute to the mounting body of evidence that synonymous mutations contribute to evolution on an unexpectedly large scale, they also show that CUB cannot be the sole explanation for why certain synonymous codons are positively or negatively selected. Were this the case, we would observe absolute usage of some codons instead of varying preferential use, which is in line with the observations of the Bailey study.

ii. mRNA and protein expression

Since synonymous mutations do not change the final encoded protein, we can look to the step preceding protein synthesis – the mRNA transcript – for clues as to how they affect fitness. Synonymous codons play a role in both the regulation of mRNA levels and their availability for ribosomal binding. For instance, a study in yeast observed that mRNA transcripts enriched with synonymous non-optimal codons had a higher decay rate than those with optimal codons (Presnyak et al. 2015). In this study, synonymous mRNA molecules of the ORF of the *HIS3* gene (which is involved in the histidine synthesis pathway) were generated to contain only optimal or non-optimal codons compared with the wild-type, which comprises 43% optimal codons. The authors found that mRNA half-life was much greater for the construct with only optimal codons

(> 60 minutes), followed by the WT (9.5 minutes), while the non-optimal codons-containing construct had a dramatically reduced half-life (2.0 minutes). The non-optimal construct was also associated with a four-fold reduction in protein expression.

Synonymous codons can also affect a transcript's thermodynamic stability, which impacts how ribosomes bind and travel along it during translation. Structures such as hairpin loops and double stranding result in a more stable secondary structure and reduced ribosomal access, which can alter translational dynamics (Kudla et al. 2009). Single-stranded sections of the transcript are easily accessed by ribosomes, but are also vulnerable to degradation by ribonucleases (RNases) (Deutscher 2006). Selection for optimal translational efficiency is thought to account for the genome-wide pattern, particularly pronounced in highly expressed genes, for reduced mRNA stability near the start codon (Gingold and Pilpel 2011).

The evolutionary constraint on mRNA structure is more important in highly expressed genes than in genes with low levels of expression (Gu et al. 2014). Deleterious synonymous mutants in *S. typhimurium* mutants were associated with decreased protein expression (Knöppel et al. 2016), which could be due to decreased translation efficiency. Another study, however, found no significant changes in mRNA levels for synonymous mutants with reduced fitness (Lind et al. 2010). In the Bailey study, the synonymous mutants exhibited a twofold increase in mRNA expression (Bailey et al. 2014). Whether this was due to increased transcription, decreased mRNA degradation, or a combination of both, is unknown. In *gtsB*, the G38G mutation decreases the predicted mRNA stability relative to the WT but A15A has no thermodynamic effect, suggesting that changes in mRNA stability are not fully responsible for the observed increase in mRNA levels. In short, these results show that synonymous mutations

clearly impact mRNA stability, which in turn affects protein output despite producing the same linear peptide sequence.

iii. Ribosomal pausing

The rate of protein synthesis may be another process affected by synonymous mutations. Synonymous mutations may be selected positively or negatively at the translational level due to their contribution to ribosomal pausing. This can occur when the translating ribosome stops at a given site on the mRNA transcript when it encounters a stretch of rare codons or an internal pause site. Phenotypically, ribosomal pausing can act in two ways, by either altering co-translational protein folding (Tsai et al. 2008) or decreasing the availability of ribosomes for translation, which can hinder growth.

In bacteria, selection seems to act against internal pause sites in the form of Shine-Dalgarno (SD) motifs (consensus sequence AGGAGG) in coding regions of genes (Chaney and Clark 2015). Normally, SD motifs are found upstream of the start codon and recruit ribosomes for translation initiation, although they can also arise by coincidence within the coding region. Their underrepresentation in the coding regions of reading frames suggests that they are here subjected to negative selection. The presence of SD-like motifs has been found to cause ribosomal pausing and therefore slower translation due to the hybridisation of the mRNA molecule and 16S rRNA, which contains the anti-SD sequence in the 3' region (Li et al. 2012), although this observation has recently been called into question (Mohammad et al. 2016). Nevertheless, translational pausing due to internal SD-like motifs can be expected to drive down the usage of codons and codon pairs that resemble SD motifs.

We can hypothesise that a synonymous change that breaks up an intragenic SD-like motif might have a positive effect on fitness, while the opposite would be true of introducing an SD-like motif in a coding region of a gene. Ribosomal pausing has also recently been shown to have a positive effect on fitness when it occurs at the start of an mRNA molecule (Frumkin et al. 2017). The "translational ramp" theory suggests that slower elongation near the 5' end is associated with lower translational cost and increased cell growth. This is due in part to the fact that more slowly translated codons prevent a downstream ribosomal traffic jam and collisions, increasing translational efficiency. Synonymous codons that encode more slowly translated amino acids might hereby have an effect on fitness.

iv. Inhibition of catabolite repression and sRNA binding

Other mechanisms that have been suggested for the effect of beneficial synonymous mutations are catabolite repression and small RNA (sRNA) binding (Bailey et al. 2014). Synonymous mutations might interfere with binding of a catabolite repression control (Crc) protein, which inhibits translation, resulting in higher expression of the gene. However, knockout of the *crc* gene in SBW25 and SBW25(A15A) and SBW(G38G) had no effect on gene expression and fitness. A second possible mechanism of inhibiting degradation is one whereby synonymous mutations interfere with the ability of sRNAs to bind to the mRNA transcript for subsequent degradation. Once again, knockout of a gene responsible for sRNA synthesis in the synonymous mutants had no effect on gene expression.

V. Thesis overview and experimental design

This project builds on the observations of the Bailey study to investigate the DFE of synonymous mutations in a gene that is targeted by selection. As we noted above, although experimental evolution provides a powerful framework to study the effects of individual mutations, empirical studies of the DFE of beneficial mutations are difficult owing to the rarity of this class of mutations and the risk of overlooking small-effect mutations. Constructing mutant strains to express a specific mutation allows us to assay the effect of many targeted mutations individually, which would be impossible in a selection experiment. Since *gtsB*, the gene identified in the Bailey study, has been shown to be the target of selection under specific conditions, this provides us with a system in which to investigate the phenotypic effect of a range of single base pair changes.

The overarching design of this experiment is simple: generate a range of *gtsB* clones in SBW25 that differ by one nucleotide and assay their fitness relative to the WT in order to isolate the phenotypic effect of individual mutations. In practice, this meant using site-directed mutagenesis to generate allelic replacements for *gtsB* (described in Chapter 2) to construct a library of synonymous and nonsynonymous point mutations at regular intervals along the gene, and subjecting each to a competition against a marked WT strain in an environment that targets *gtsB* for selection. From the resultant range of fitness values, we were able to find the optimal parameters that describe the shape of the distribution of beneficial mutations. Crucially, this method allowed us to determine the fitness effects of many synonymous mutations, in order to address the gap in our collective knowledge regarding their contribution to adaptive evolution.

Using our relative fitness estimates for dozens of synonymous and nonsynonymous mutations, we were also able to perform *in silico* analyses to determine whether a mechanism of action could be identified for the fitness effects of synonymous mutations. Since A15A and

G38G were both located near the start of the gene, we asked whether beneficial mutations are more likely to occur in this location – especially considering that genetic architecture at the 5' end of genes appears to be under stronger evolutionary constraint than the rest of the genes (Tuller et al. 2010; Bentele et al. 2013; Frumkin et al. 2017). Furthermore, we tested whether any of the CUB, mRNA stability, or internal ribosomal pausing sites could explain the variations in fitness. Ultimately, by allowing us to characterise the DFE of synonymous mutations in a gene under selection, this project contributed to characterising the role played by synonymous mutations in adaptive evolution.

CHAPTER 2: METHODS

I. Site-directed mutagenesis of *gtsB*

Site-directed mutagenesis of *gtsB* was accomplished by cloning *gtsB* alleles with a single mutation into a vector, generating an *E. coli* library, followed by performing allelic replacements in SBW25 (outlined in Fig. 5). Primers were designed to introduce mutations at 112 sites in the *gtsB* gene (*PFLU4845*), corresponding to 50 codons spanning the gene (every 10th codon) and including clusters around the previously identified adaptive mutations (codons 10, 15 and 38) (Bailey et al. 2014). Sequences from 715 base pairs upstream to 173 base pairs downstream of *gtsB* were amplified as two polymerase chain reaction (PCR) fragments, one of which contained a threefold degenerate polymorphism introduced by a mutagenic primer. Each mutagenic primer was paired with a conserved reverse primer (R3_gtsB_R) to amplify *gtsB* sequences downstream of the mutated site. Sequences upstream of the mutated site were amplified using a conserved forward primer (F3_gtsB_R) paired with a site-specific primer annealing adjacent to the mutated locus. BsaI recognition sequences were included in each primer to enable seamless ligation between the *gtsB* PCR products and an allelic replacement vector using Golden Gate assembly (Engler et al. 2008). Since the BsaI endonuclease cuts distal from its recognition sequences, both BsaI and DNA ligase could be included with the allelic replacement vector and purified PCR products in a single cloning reaction. The cloning reactions were incubated at 37°C for two hours, 50°C for five minutes, and 80°C for five minutes, then held at 10°C and transformed into *Escherichia coli* DH5⟨⌊pir following the Inoue method for chemical transformation (Sambrook et al. 1989) with selection on ampicillin.

Transformations yielded libraries of *E. coli* strains that were used to introduce mutations (corresponding to the threefold degenerate sites of the mutagenic primers) into *P. fluorescens*. Recombination of each mutant *gtsB* allele into the chromosome was selected for by two-step allelic replacement, with selection for tetracycline resistance followed by selection for sucrose resistance as previously described (Bailey et al. 2014). To increase the efficiency of replacement, we used an SBW25 recipient strain in which the native *gtsB* was replaced by *lacZ*, allowing us to use blue-white screening on LB 5% sucrose X-gal agar to identity recombinants in which *lacZ* was replaced by the vector-encoded mutant *gtsB.* The sucrose-resistant white colonies were used as PCR templates for amplification of the *gtsB* locus using an M13F-tagged primer, followed by sequencing performed by the McGill University and Genome Quebec Innovation Centre. Coverage was an average of 960 base pairs downstream of the start codon, covering the 909-nucleotide length of *gtsB*; 152 unique point mutants were identified.

II. Fitness estimates

Competitions were performed as outlined in Lenski et al (1991) (Lenski et al. 1991) on four to six technical replicates of 111 mutant strains, removed from storage in glycerol at -80°C and grown overnight at 28°C on LB agar, as was the SBW25 *lacZ* competitor. Single colonies were inoculated into 2 mL LB broth for overnight shaking incubation at 28°C. Cultures were then transferred into minimal glucose media for a 24-hour acclimation period at 28°C, after which each mutant strain was mixed in a 1:1 ratio with SBW25 *lacZ* into 2 mL of minimal media supplemented with glucose. Samples from the mixed culture were frozen in 20% glycerol after one- and 24-hour periods, which we refer to as initial and final time points respectively. The frequencies of the two strains at each time point were obtained by counting at least 30 colonies

of each after dilution and spreading on minimal glucose agar with X-gal. Identification of mutant and competitor colonies was possible since the mutants, lacking *lacZ*, produced white colonies, while SBW25 *lacZ* colonies were blue. Relative fitness was calculated using the equation $\omega = (f_{final}/f_{initial})^{(1/generations)}$, where $f_{initial}$ and f_{final} are the ratios of the frequency of mutant strain to the frequency of SBW25 *lacZ* strains before and after the competition. The number of generations was estimated to be approximately 6.7 or 13.2, depending on the dilution factor. To verify that *lacZ* has no effect on fitness in SBW25, it was competed against the WT SBW25 in each fitness assay. Values from these assays in glucose were combined, resulting in a relative fitness estimate of 1.005 ± 0.0007 S.E.M. Competitions for a subset of mutants were also performed in minimal media supplemented with either mannitol (N = 25) or succinate (N = 23). The *lacZ* strain competed against the WT showed a small fitness cost in both mannitol and succinate, in the order of 0.018 and 0.006. Relative fitness values were therefore adjusted to reflect a WT value of 1 by subtracting 0.018 and 0.006 from each estimate in mannitol and succinate respectively.

III. Media

Escherichia coli was grown on Luria-Bertani (LB), X-gal sucrose, or tetracycline media. *Pseudomonas fluorescens* SBW25 was grown on LB or X-gal minimal salts media (48mM Na_2HPO_4, 22mM KH_2PO_4, 9mM NaCl, 19mM NH_4Cl, 2mM $MgSO_4$, 0.1mM $CaCl_2$) with glucose (53 ⌠M), succinate (80 ⌠M), or mannitol (53 ⌠M) as indicated. Media were supplemented with 5-bromo-4-chloro-3-indolyl-®-D-galactopyranoside (X-gal) at 40 ⌠g/ml. Antibiotics were used at the following concentrations: 100 ⌠g/ml nitrofurantoin (Nf), 100 ⌠g/ml ampicillin (Ap), 10 ⌠g/ml tetracycline (Tc).

IV. Estimating codon preference and mRNA stability

Following the procedure detailed in Bailey et al (2014), we estimated the change in codon bias attributable to each synonymous mutation by the codon adaptation index (CAI) value of the mutant to the WT using SBW25 ribosomal protein genes as a reference of highly expressed genes. The 'cai' function in the 'seqinr' package in R was used to calculate change in CAI at each site. tRNA adaptation index (tAI) values were calculated by inputting tRNA gene copy numbers as an experimentally validated proxy for tRNA expression (Tuller et al. 2010) in the $stAI_{calc}$ interface. Consistent with previous work (Kudla et al. 2009), we predicted the most probable folding energy of 42-nucleotide windows centred on each mutation using the Quikfold application of the 'mfold' server (Zuker 2003). This window size was chosen was it corresponds to the approximate outer limits of a ribosomal subunit binding site on the transcript (Takanami and Zubay 1964).

V. Statistical analyses

All statistical analyses were conducted in R Studio (version 1.0.136; https://www.rstudio.com). We tested whether the DFEs of synonymous and nonsynonymous mutations were drawn from the same distribution using the Kolmogorov-Smirnov (K-S) test on bootstrapped data. The K-S test is suitable for analysing our dataset since it assumes a continuous distribution and is nonparametric (neither the synonymous nor the nonsynonymous samples were normal (Shapiro-Wilk test, $P = 6 \times 10^{-9}$ and 6×10^{-8} respectively)). In order to calculate a P-value, we generated a null distribution of the K-S D-statistic by bootstrapping the nonsynonymous sample. To do so, we performed 10,000 permutations of the following:

sampling the nonsynonymous sample, with replacement, to generate a simulated sample the same size as the synonymous sample and performing a K-S test comparing the bootstrapped sample with the observed nonsynonymous sample. From the resultant null distribution of D-statistics, we computed a one-sided P-value by calculating the proportion of bootstrapped D-statistics that were greater than or equal to the true D-statistic (which compared the observed synonymous and nonsynonymous samples).

We sought to detect a mechanism responsible for the variation in fitness effects, to which end we fit linear models to the data with the following variables as main effects: type of mutation (synonymous/nonsynonymous), transition/transversion, amino acid location along the gene, change in CAI, change in tAI, and change in ΔG. Relative fitness values were weighted by the inverse of their standard error such that more precise estimates weighed more heavily. P-values for the linear models were obtained by performing permutations (N = 10,000), a method that also addressed the non-normality of the residuals. A best model, combining the variables found to be significant, was built using performing a permutation (N = 10,000) on the residuals (Still and White 1981) to generate a null distribution of the F-statistic and determine P-values. Our threshold for significance was established at $\alpha = 0.05$.

CHAPTER 3: RESULTS AND DISCUSSION

I. Library of *gtsB* point mutants

As described in Chapter 1, although purifying selection on synonymous sites has been demonstrated in various systems (Lind et al. 2010; Cuevas et al. 2012; Lawrie et al. 2013), reports of beneficial synonymous mutations are rare (She and Jarosz 2018). The discovery of two independently and spontaneously arising beneficial synonymous mutations in a *P. fluorescens* population provided further evidence that synonymous mutations can be adaptive, although in general, their frequency of occurrence is unknown. These mutations arose in *gtsB*, which is the second gene in an operon encoding an ATP-binding cassette transporter that transports glucose across the inner membrane. Conferring fitness increases of 6-9%, these two mutations' effects are on par with those of a nonsynonymous mutation that also arose independently in *gtsB* during the experiment. The mutations to *gtsB* had no effect on fitness in alternate carbon sources, confirming that *gtsB* is a target of selection specifically during growth on glucose (Bailey et al. 2014). This discovery provided unequivocal evidence that synonymous mutations can increase in frequency via adaptation rather than via genetic drift or hitchhiking with beneficial nonsynonymous mutations. In order to apply the theory underlying adaptation, outlined in Chapter 1, to synonymous mutations, we had to construct a library of *gtsB* point mutants and measure each mutation's fitness effect.

Following the method detailed in Chapter 2, we used site-directed mutagenesis to construct 40 synonymous and 71 nonsynonymous sequence-confirmed substitutions at 114 loci spanning the length of *gtsB* in *P. fluorescens* SBW25. Relative fitness for each mutant was measured in four to six technical replicate competitions against a marked ancestor performed in minimal

media supplemented with 53 ⌈M glucose for 24 hours at 28°C. Fitness estimates of the competitor show that it has neutral relative fitness (mean ⌉ = 1.004, ± 0.007 S.E.M.), while a *gtsB* knockout strain (⊗*gtsB*) has greatly decreased fitness in glucose (0.716 ± 0.014), although not in mannitol or succinate media (1.016 ± 0.025 and 1.019 ± 0.020 respectively). Relative fitness estimates ranged from 0.70 to 1.08 (median = 1.01) for synonymous mutations and 0.72 to 1.05 (median = 0.99) for nonsynonymous mutations (Fig. 1). Fitness effects observed in glucose did not carry over to mannitol- or succinate-supplemented minimal media (Fig. 4).

The focus of this study was to compare the effects of synonymous and nonsynonymous mutations in an otherwise intact protein; for this reason, six nonsense mutations from our original library of *gtsB* mutants were omitted from our analysis, since their strongly deleterious effect is likely due to the production of a truncated protein. Consistent with the theory that the wild-type (WT) gene is already well-adapted and any change to its sequence is more likely to disrupt function than improve it (Gillespie 1984), a prediction supported by empirical evidence (Kassen and Bataillon 2006), the DFE of nonsynonymous mutations comprises mainly deleterious mutations with a small proportion of neutral and beneficial mutations (43%). Comparatively, the DFE of synonymous mutations reveals a larger proportion (68%) of beneficial effects.

II. Comparison and characterisation of distributions of fitness effects

As detailed in Table 1, we compared both the entire dataset and only beneficial mutations using a permutation of the K-S test. The distributions of the former differ (P = 0.0002), with the nonsynonymous sample including a highly deleterious secondary peak (Fig. 1). Strikingly, however, the DFEs did not differ when we considered only beneficial mutations (P = 0.59) (Fig.

2), allowing us to pool them for subsequent analyses. This is a novel observation, although similar effect sizes in beneficial synonymous and nonsynonymous mutations have been previously described (She and Jarosz 2018). Thus, we sought to characterise the shape of the distribution of adaptive mutations as informed by theoretical models.

Theory predicts that adaptation is driven by a pattern of diminishing returns – i.e., that the probability of occurrence of beneficial mutations is inversely proportional to the magnitude of their effect (Orr 2005). Ronald A. Fisher's geometric model of adaptation was the first to predict that small-effect beneficial mutations were more frequent than large-effect beneficial ones, and that the former were therefore the basic unit of adaptation (Fisher 1930). Shifting the focus from phenotypes, as in Fisher's model, to mutations, John Gillespie proposed that the DFE of non-beneficial mutations, which are purged by selection, is irrelevant to modeling adaptive evolution (Gillespie 1984). He argued that as beneficial mutations represent rare events above a high threshold (the WT enjoying high fitness), they can be considered in the framework of extreme value theory (EVT) described by a Generalised Pareto Distribution (GPD) (Gillespie 1984; Joyce et al. 2008).

The GPD family encompasses three domains of attraction that are determined by the value of the shape parameter |, also referred to as the tail index, which specifies the weight of the tail (Beisel et al. 2007; Joyce et al. 2008). Currently, the nature of the distribution of beneficial fitness effects is still under investigation – and the DFE of beneficial synonymous mutations is entirely unknown. Although an exponential DFE, where | = 0 (belonging to the Gumbel domain) has received theoretical and empirical support (Rozen et al. 2002; Orr 2003), the Weibull domain, with a truncated tail (| < 0), is emerging as a candidate as well, with experimental

support for an upper bound on fitness effects (Rokyta et al. 2008). Finally, the heavy-tailed Fréchet domain ($\xi > 0$) has not received empirical validation.

Since beneficial mutations are rare, empirical tests of their DFE are few (Eyre-Walker and Keightley 2007; Hunt et al. 2014). Selection experiments, in which the first mutations to fix are isolated, produce biased samples as these mutations must escape both drift and clonal interference. By constructing a library of point mutants, we generated a larger, unbiased sample from which we characterised the shape of the right-hand tail of the DFE, thereby including all beneficial-effect mutations ($\lambda > 1$) irrespective of formal tests for significance. This tail takes the form of a GPD, which, as detailed above, comprises the Gumbel, Fréchet, and Weibull domains of attraction. Our analysis thus focused on defining the shape of the tail of the beneficial DFE. As we will detail next, we tested the null hypothesis that the DFE of beneficial mutations has an exponential tail by restricting ξ to 0 against the alternate hypothesis that a distribution with an unrestricted shape parameter provided a better fit.

Beisel et al (2007) provide a likelihood ratio framework tailored to test the null hypothesis that a distribution of beneficial fitness effects is exponential (a special case of the Gumbel domain). To that end, we first transformed the data from relative fitness values (λ) to selection coefficients (s) by subtracting 1 from all relative fitness estimates ($\lambda = 1 + s$). Next, we further transformed the data with a simple shift such that each s value became relative to the smallest observed selection coefficient rather than relative to the WT. The establishment of this new threshold, to which ξ is robust, is necessary to avoid misinterpreting the distribution of unobserved values between the WT threshold and the smallest selection coefficient, although it comes at the cost of a degree of freedom.

Following this shift, we computed bootstrapped maximum likelihood estimates for the model with the restriction | = 0 (which corresponds to the exponential distribution). We also estimated the maximum likelihood of the alternative model, where | is not restricted, by calculating the optimal | value using the GenSA package in R (Xiang et al. 2013). Since the exponential distribution, as a special case of the Gumbel domain, is nested within the GPD, we can compare the restricted model with the unrestricted |-value model using a likelihood ratio test (LRT). To determine significance, we calculated the P-value as the proportion of bootstrapped likelihood ratio tests (N = 10,000) that were greater than or equal to the true LRT.

We find that an optimal | value of -0.37, belonging to the Weibull domain, describes the distribution better than the exponential form of the GPD (P = 0.0077). In biological terms, the truncated tail unique to the Weibull domain translates to an upper bound on fitness effects among beneficial mutations available to selection; this result adds to mounting empirical evidence favouring the Weibull domain, suggesting that beneficial mutations facing selection arise in a landscape involving many small-effect beneficial mutations (Rokyta et al. 2008; Betancourt 2009; Schoustra et al. 2009; Bataillon et al. 2011). Furthermore, it is consistent with the biologically realistic interpretation of Fisher's model, which assumes a single phenotypic optimum. Critically, though, we show that the beneficial DFE of synonymous mutations also belongs to the Weibull domain, suggesting that they are an overlooked source of adaptive potential.

III. Molecular mechanisms responsible for fitness variation

In order to elucidate a mechanistic explanation for the fitness effects of both synonymous and nonsynonymous mutations, we investigated, independently, the main effect of several

variables: mutation type, transition/transversion, amino acid location along the gene, change in CAI, change in tAI, and change in mRNA free energy (ΔG). As shown in Table 2, only the models testing the main effects of mutation type and change in mRNA free energy were significant. Transitions occur when the mutant nucleotide belongs to the same class as the WT – either a purine, to which the double-ringed adenine and guanine belong, or a pyrimidine, which includes the single-ringed cytosine and thymine. The model including transition v. transversion as the dependent variable was not significant. Similarly, models including change in CAI and change in tAI was were not significant. Moreover, we note that we did not identify any consistent fitness effects – whether beneficial or deleterious – based on internal SD-like sites. Only the main effects of mutation type and change in predicted mRNA transcript stability were significant and combined into a 'best' model. We detail our findings below.

Synonymous mutations can play a role in altering translational dynamics by encoding a more or less preferred codon (Komar et al. 1999; Cortazzo et al. 2002). Codon optimality can be estimated by assigning a codon adaptation index (CAI) value based on the relative synonymous codon use in a set of highly expressed ribosomal protein genes (Sharp and Li 1987). Evidence of gene architecture near the 5' terminus being important for proper initiation of transcription and translation (Tuller et al. 2010; Navon and Pilpel 2011) is supported by an analysis of over 5000 genes in another Pseudomonad, *P. aeruginosa*, showing an overall trend for the first 50 codons of genes to be enriched in suboptimal codons (Tuller et al. 2010). We find no evidence of such a CAI ramp in *gtsB* ($R^2 = 0.0057$, $P = 0.21$), and change in CAI does not explain the variation in fitness effects in the synonymous sample (permutation of residuals, $P = 0.41$) (Fig. 2C). Although CAI is positively correlated to the relative abundance of cognate tRNAs (Ikemura 1985), a model examining the effect of tRNA adaptation index (tAI) values on fitness was

similarly not significant; tAI, unlike CAI, further accounts for species-specific wobble weights ($P = 0.53$). In our dataset, each mutation is a single-codon change, which represents such a minute change in CAI when considering the entire gene that there is no measurable effect; this interpretation is supported by previous work suggesting that the effect of changing codon preference on gene expression is only detectable when substantially manipulated (Kudla et al. 2009). Furthermore, previous studies have shown that the position of a synonymous mutation in a gene has no impact on the magnitude of its effect (She and Jarosz 2018).

As described in the Introduction, synonymous mutations may also affect translation speed and hence protein folding, as well as the rate of mRNA transcript degradation, by altering the stability of the mRNA transcript (Marin 2008; Saunders and Deane 2010). We obtained predicted transcript stability values using the 'mfold' RNA folding server for a 42-nucleotide window centred on each mutation (the approximate limits of nucleotide interaction) (Kudla et al. 2009). After analysing possible main effects, we arrived at a model combining the statistically significant main effects of mutation type and change in mRNA stability. As noted previously, although tested, the interaction of these two effects was not significant. The model itself was significant, but explained very little of the observed variation (adj. $R^2 = 0.19$, $P = 5.7 \times 10^{-5}$). Although the main effect of mRNA stability was significant, we did not find a biologically compelling link between mRNA secondary structure and fitness for synonymous mutations, as we will detail next (Fig. 3A).

Since we did not know the relationship between mRNA stability and fitness, we considered that fitting a non-linear model, which makes no assumptions about the nature of the relationship, was most appropriate. A second-order polynomial relationship best described the fitness variation observed (permutation of residuals, $P = 0.0014$). We then proceeded to compute

coefficients for the shape of the curve of the polynomial for the synonymous and nonsynonymous samples separately. The nonsynonymous sample exhibits a curve driven by an association of highly deleterious, less-fit mutations. This effect may potentially be explained by the trade-off between the increased availability for translation and increased degradation of less-stable mRNA (Chamary and Hurst 2005). However, since the fitness effects of nonsynonymous mutations were not the focus of this study, we did not pursue this further. When fitting a curve with the coefficients specific to the synonymous sample, we found that although statistically significant, the relationship was not biologically so since the effect size was minute (yielding a nearly flat slope). For this reason, we only included the slope associated with the nonsynonymous group in Figure 3A. Finally, while strong secondary structure at the 5' end of the open reading frame has been associated with decreased cost of protein translation (Frumkin et al. 2017), adding a position term does not improve the fit of this model.

Our results clearly show that beneficial synonymous mutations are much more common than previously thought, and that synonymous mutations have, on the whole, a neutral to beneficial effect. Additionally, the DFE of beneficial synonymous mutations is indistinguishable from that of nonsynonymous mutations, suggesting that they have similar effect sizes, concordant with a recent nucleotide-level genome-wide analysis of *S. cerevisiae*, which found that nonsynonymous mutations play in a role in diversifying protein function, whereas synonymous mutations modulate protein expression (She and Jarosz 2018). We identify a pattern of diminishing returns that produces a bounded distribution belonging to the Weibull domain of the GPD for both synonymous and nonsynonymous mutations. A single mechanism by which synonymous mutations increase fitness remains elusive. Nevertheless, the unexpectedly large

proportion of beneficial synonymous mutations identified suggests that they are an underestimated source of adaptive potential.

CHAPTER 4: CONCLUSIONS

I. Summary of findings

The role played by synonymous mutations in adaptive evolution has been largely overlooked. Although there is mounting evidence that synonymous sites can be under both purifying and positive selection across all domains of life, we have yet to arrive at a general understanding of the fitness effects of synonymous mutations. Through this Master's project, I sought to address this knowledge gap by using experimental evolution to generate DFEs for synonymous and nonsynonymous mutations in a gene known to be under selection. To do so, I assayed the relative fitness of a library of point mutants for the *gtsB* gene in *P. fluorescens* in minimal media supplemented with glucose, an environment where *gtsB* is a target of selection.

Overall, I found that the DFEs of synonymous and nonsynonymous mutations differ owing to the presence of nonsynonymous mutations that have strongly deleterious effects, many of which are premature stop codons that presumably lead to a truncated protein. Strikingly, however, given the common assumption that synonymous mutations are selectively neutral, the DFE of beneficial mutations is not significantly different between the two groups. Bolstered by recent work in yeast showing similar effect sizes for both synonymous and nonsynonymous mutations (She and Jarosz 2018), this result constitutes convincing experimental evidence for the inclusion of synonymous mutations in models of adaptive evolution, as I will detail later.

The development of a theory of adaptive evolution has involved exchanges between theoretical and empirical work. However, the experimental study of beneficial mutations – the basis of adaptive evolution – is limited by two factors: their rarity and the fact that most have a small effect size. First, since the WT is already well adapted, fitness-enhancing mutations are

infrequent and most mutations are deleterious and purged by selection. Secondly, theory predicts that most beneficial mutations have a small effect size, which makes them vulnerable to being lost due to drift or clonal interference. Experimental evolution approaches to inferring the DFE of beneficial mutations thus suffer from small sample sizes and bias. In a selection experiment, for instance, a population is challenged by a new environment and mutations are isolated when detectable – at or near fixation. In this approach, only spontaneously arising mutations ever get the chance to be tested by selection, and only those that escape both drift and possibly competition from other beneficial mutations can be detected.

By constructing and directly phenotyping individual mutant strains, however, I was able to assay many targeted mutations, resulting in a larger, unbiased sample. I constructed a library of point mutations in *gtsB* via site-directed mutagenesis and used this to estimate fitness for over a hundred genetically unique strains in a single gene known to be under selection. Combined into DFEs, these fitness estimates provided the raw data to which I could apply the theoretical framework of adaptive evolution. Most importantly, the gene I manipulated is biologically relevant, being one in which multiple mutations arose spontaneously and conferred an adaptive advantage specifically during growth on glucose – but not on other sugars – demonstrating that it is a target of selection in an environment with limiting glucose. My study system thus allowed me to directly measure the effect of individual, targeted mutations on fitness in an environment where selection favours the ability to grow on limited glucose, and to identify even small effect sizes.

My comprehensive evaluation of the fitness effects of point mutations throughout *gtsB* enabled me to characterise the shape of the DFE of beneficial mutations, prior to the effects of drift and selection. Models of adaptation treat the fitness effects of beneficial mutations as draws

from the extreme right-hand tail of a broader, often unknown, distribution of fitness effects. Viewed in this way, the DFE among beneficial mutations that are available to selection can be described using extreme value theory (Gillespie 1984). I found that my observations were best described by the Weibull domain of attraction of the Generalised Pareto Distribution family, similar to some previous predictions and estimates (Martin and Lenormand 2008; Rokyta et al. 2008), although traditionally, the exponential distribution, which is a special case of the Gumbel domain, has been favoured (Orr 2003, 2006). Biologically, a DFE belonging to the Weibull domain means that there exists a maximum fitness – a phenotypic optimum. In Maynard Smith's fitness landscape framework, this corresponds to the point where a population has evolved to a stage where all possible changes result in a less-fit genotype (Maynard Smith 1970).

My *in silico* investigations did not reveal a single mechanism responsible for the variation in synonymous mutations' effect on fitness, although there are promising avenues of study. The original Bailey study showed that the two beneficial synonymous mutations increased fitness through increased gene expression via increased transcript abundance, an observation that has been repeated in *M. extorquens* and *S. cerevisiae* (although both increased expression via increased protein abundance) (Agashe et al. 2016; She and Jarosz 2018). Of note, all three studies involve small sample sizes – involving just one to four beneficial synonymous mutations. Future work could exploit the larger sample size contained in my library of constructs to explore whether synonymous mutations are beneficial through increased transcription or translation.

My results also raise further questions: for instance, given that my findings were in a gene that is essential for growth in a specific environment and therefore a target of selection, is the resulting DFE specific to similarly important genes? Moreover, do the similarities in DFEs of synonymous and nonsynonymous mutations carry over to genes that are not under selection?

Finally, just as the effect of a nonsynonymous mutation is context-dependent – that is, the new mutant amino acid has an effect in the context of the wild-type sequence by altering protein conformation – could synonymous mutations also be highly context-dependent, as has been put forward by others (Agashe et al. 2016)? Larger datasets of single synonymous variants, coupled with molecular work focusing on the effect of synonymous mutations on the transcriptional and translational processes, are therefore necessary to arrive at a mechanistic explanation.

II. Implications for evolutionary biology

i. Adaptive walks

My results can be considered in the context of adaptive landscapes and the role played by synonymous mutations on the path to a fitness peak. Adaptive landscapes, as proposed by Kauffman and Levin, are a space where "entities" – that is, genotypes – live in an adaptive space and move upwards, or at least sideways, to a more fit peak through a series of one-step variants (Kauffman and Levin 1987). And so, traditionally, adaptation has been studied as a series of sequential genetic changes that occur when a population is challenged by a novel environment, resulting in an adaptive walk. Assuming a low mutation rate, which means that double and triple mutants are highly unlikely and thus can be ignored, beneficial single mutations have the potential to bring the population one step closer to a local optimum. However, due to the architecture of genes, only a discrete number of more-fit sequential substitutions is available to any given genotype (Orr 2002).

Synonymous mutations, if assumed to be neutral, result in a flatter fitness landscape. Where only amino acid changes are considered, a population can theoretically move through several synonymous codons without altering protein function or fitness before arriving at a different

amino acid that confers a fitness advantage. Conceptually, this can be viewed as travelling without moving: by conserving the amino acid sequence, fitness is maintained while travelling to a different part of the protein adaptive landscape. Neutral synonymous intermediates are thus employed to allow a protein to explore different adaptive spaces without resulting in either reduced or improved fitness (Cambray and Mazel 2008). However, this is unlikely to occur in a large population such as that involved in my study system, where the effect of drift is weak, and neutral and even mildly deleterious mutations are thus unlikely to rise to fixation.

My observations suggest that the assumption that synonymous mutations are neutral steps along an adaptive walk is incorrect. Given that synonymous mutations in a gene facing selection may not be neutral, we can in fact imagine a much more rugged landscape, one in which every nucleotide-level change leading to a local optimum could be synonymous just as well as nonsynonymous. Concordant with my findings, recent work in HIV has linked synonymous mutations to both small and, at important loci, high fitness costs, corresponding to dips and valleys in the fitness landscape (Zanini et al. 2017). These results are supported by recent computational work in which a landscape that includes synonymous mutations has more peaks than one that only considers amino acid-level changes (Fragata et al. 2018). As a result, a population may then stall at a local optimum with intermediate fitness created by variation in the effects of synonymous mutations.

ii. Mutation rate calculation

Inferring patterns of selection from SNP data is a well-used method to study evolutionary processes. The ratio between the rate of substitution at nonsynonymous sites (dN) to the rate of substitution at synonymous sites (dS) has been commonly used to identify loci targeted by

natural selection, with the assumption that the synonymous mutations, being neutral, effectively reflect the background mutation rate (Nielsen 2005). If most nonsynonymous mutations are deleterious, then dN/dS < 1, indicating negative selection. Conversely, a dN/dS ratio greater than 1 is interpreted as a signal of positive selection. This method has been in use since the 1980s (Hughes and Nei 1988; Stahl and Bishop 2000; Ford 2002; Kryazhimskiy and Plotkin 2008; Good et al. 2017), despite growing evidence that the assumption of universal neutrality for synonymous mutations is erroneous, as we detailed in Chapter 1.

My observation of both neutral and beneficial effects of synonymous mutations suggests that, if the gene in question is a target of selection, estimates of dN/dS may be inaccurate. If some synonymous mutations are under positive selection in a gene, as we observed in *gtsB*, we risk overlooking that region if using dN/dS to assess the strength of selection. Even realistic evolution models take into account only codon and transition/transversion biases, ignoring the adaptive potential of synonymous substitutions. Considering that beneficial synonymous mutations have a comparable DFE to beneficial nonsynonymous mutations, estimates of neutral mutation rates should avoid assuming the selective neutrality of synonymous mutations if the gene in question is targeted by selection.

III. Closing remarks

By showing that adaptive synonymous mutations are not an anomalous occurrence, but in fact follow the same distribution as nonsynonymous mutations in a gene targeted by selection, my Master's work adds to the mounting body of evidence showing that synonymous mutations may face the same selective pressures as nonsynonymous changes. Although the realisation that synonymous changes may play a much larger role in adaptive evolution than previously thought

is important in itself, this work also leads us to reconsider further the basis of common theories of adaptive evolution that have been built on the assumption of 'silent' synonymous changes. Thus, the trajectory of adaptive walks and the shape of fitness landscapes, as well as mutation rate calculations through SNP analyses, should all consider the possibility of beneficial synonymous mutations. If a gene is the target of selection, we cannot dismiss synonymous mutations with the assumption that theirs is a silent role: my research shows that they may in fact play a clear, and heretofore underestimated, role in adaptive evolution.

FIGURES

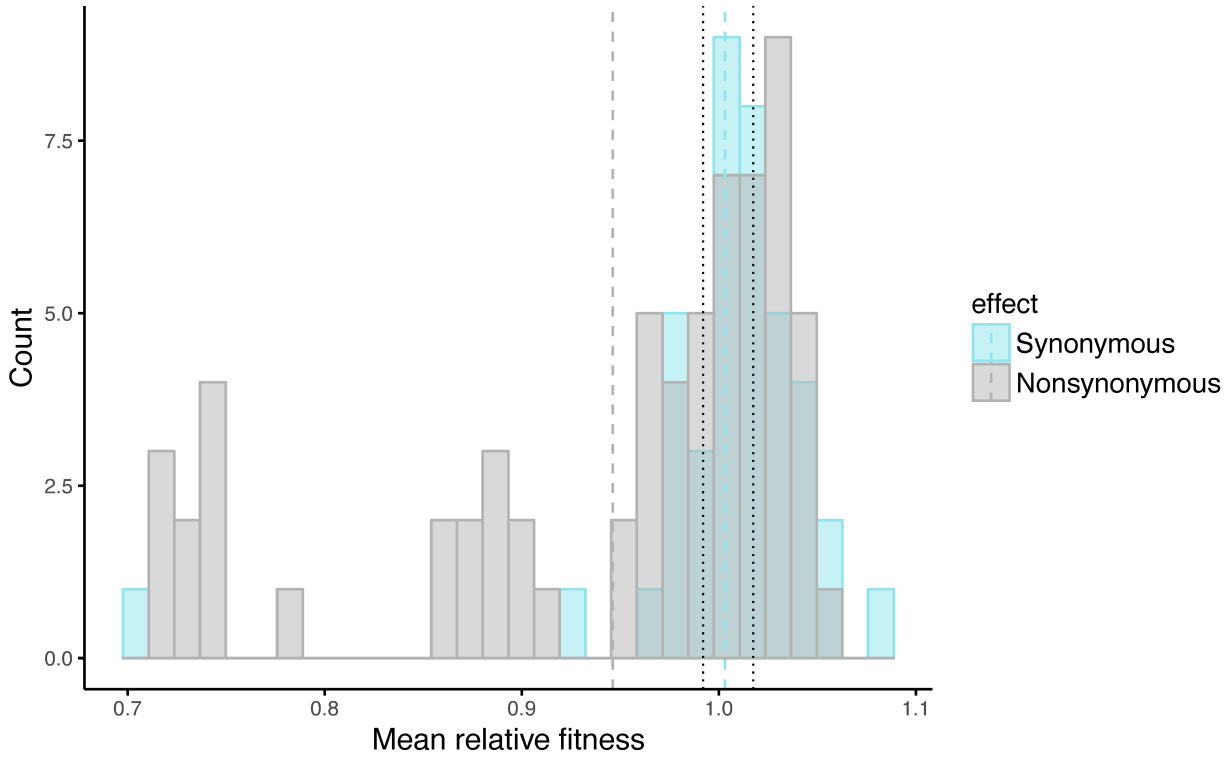


Fig. 1. Distribution of fitness effects of the *gtsB* point mutant library. Gene variants of *gtsB* were synthesised using site-directed mutagenesis and expressed in *P. fluorescens* SBW25 through allelic replacements and competed against SBW25 *lacZ* in low glucose media. Relative fitness values are estimated from the change in the ratio of mutant to ancestor frequency after a 24-hour competition. Dashed lines indicate the mean of each sample (synonymous N = 40; nonsynonymous N = 65); dotted lines indicate the 95% confidence interval of the WT competitions (N = 17).

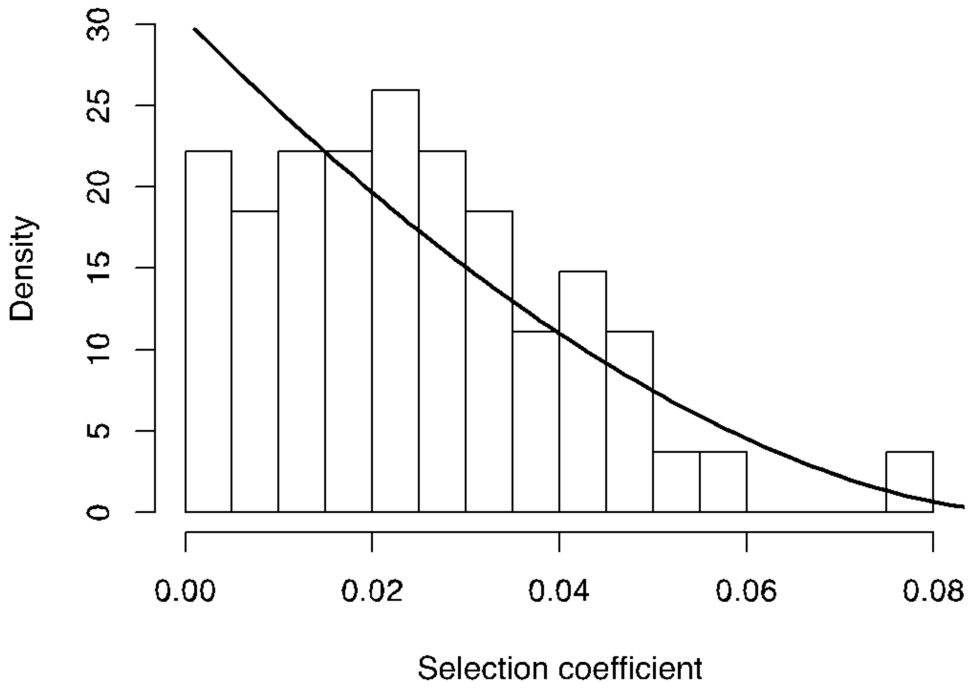
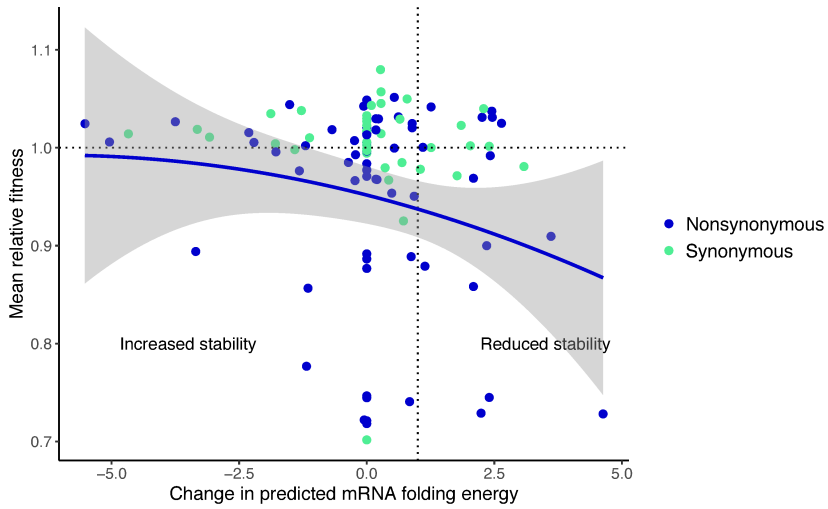
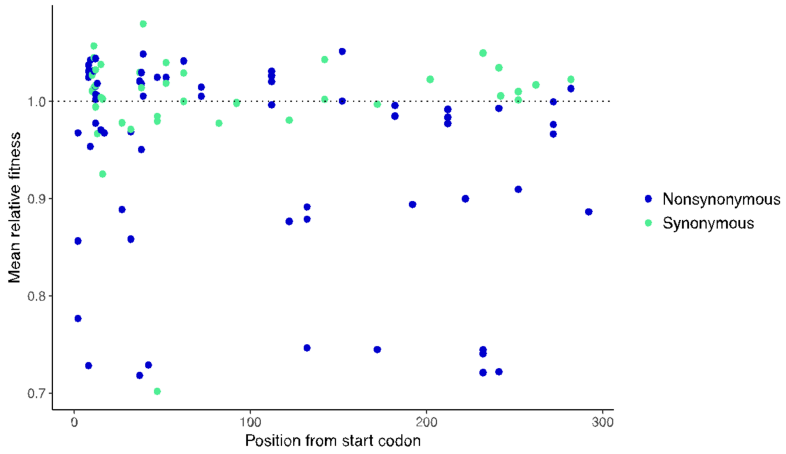


Fig. 2. Distribution of fitness effects of beneficial mutations. The DFE of the pooled synonymous and nonsynonymous samples (N = 54) is fit by a tail index of -0.35, which belongs to the right-truncated Weibull domain of attraction of the Generalised Pareto Distribution.

A


Mean relative fitness
1.1
1.0
0.9
0.8
0.7
Increased stability
Reduced stability
−5.0
−2.5
0.0
2.5
5.0
Change in predicted mRNA folding energy
Nonsynonymous
Synonymous

B


Mean relative fitness
1.0
0.9
0.8
0.7
0
100
200
300
Position from start codon
Nonsynonymous
Synonymous

C

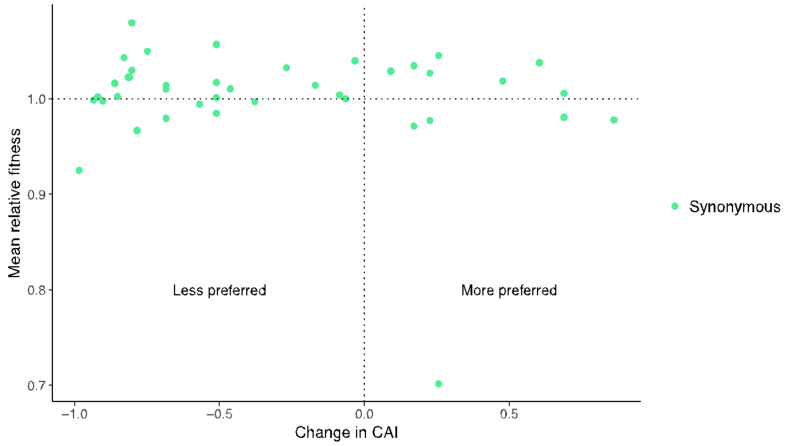

Mean relative fitness
1.0
0.9
0.8
0.7
Less preferred
More preferred
−1.0
−0.5
0.0
0.5
Change in CAI
Synonymous

Fig. 3. *In silico* analyses. (A) The association between decreased predicted mRNA transcript stability of nonsynonymous mutations (blue) and decreased fitness was best described by a polynomial model (P = 0.03); shaded area indicates 95% confidence interval of the model. Change in mRNA stability had no effect on the fitness effects of synonymous mutations (green). (B) There was no effect of a mutation's position within the gene for either synonymous or nonsynonymous mutations. (C) Whether a mutation resulted in a more- or less-preferred codon had no impact on the observed fitness effect (synonymous mutations only).

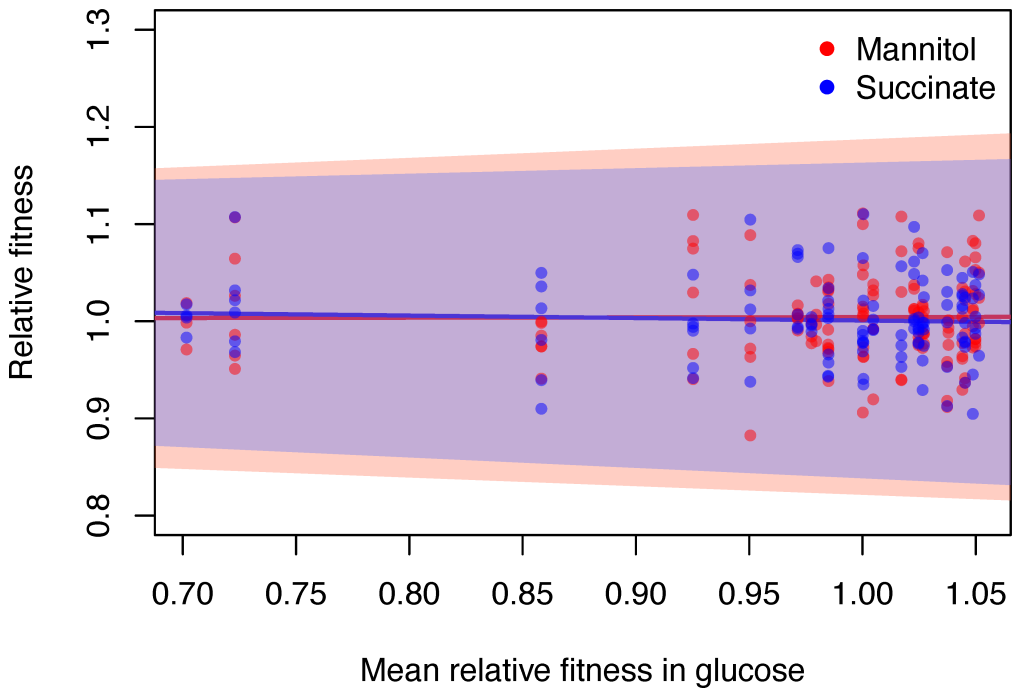


Fig. 4. Mutational effects in alternate carbon sources. The relative fitness values of a subset of synonymous and nonsynonymous mutant strains (N = 20) in minimal media with mannitol (red) or succinate (blue) are plotted as a function of their mean relative fitness in glucose. Shaded areas indicate the 95% confidence interval for the slope of the linear model for each environment.

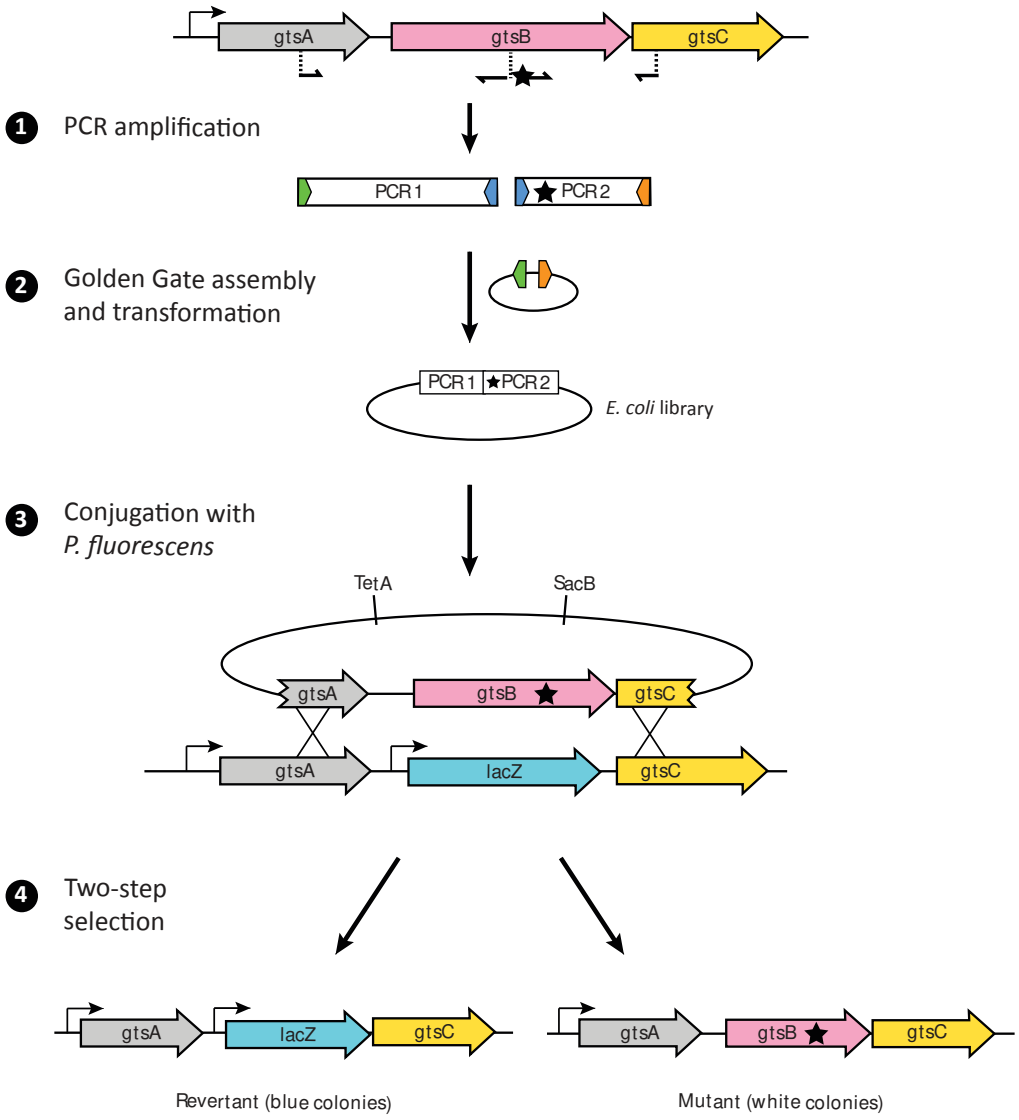


Fig. 5. Diagram illustrating site-directed mutagenesis of *gtsB* using Golden Gate cloning.

TABLES

Table 1: Bootstrapped analysis of the Kolmogorov-Smirnov statistics for the DFEs. Nonsense mutations were omitted from our analysis.

Distributions tested	**Mutations included**	**P**
Synonymous v. nonsynonymous (glucose)	All	0.0006
Synonymous v. nonsynonymous (glucose)	Beneficial	0.5868
Glucose v. mannitol	Subset of 20 strains	0.0138
Glucose v. succinate	Subset of 20 strains	0.009
Mannitol v. succinate	Subset of 20 strains	0.2749

Table 2: Permutation tests on linear models investigating the effect of explanatory variables for fitness differences. Nonsense mutations were omitted from our analysis.

Term	**P**
Mutation type	0.0018
Transition v. transversion	0.4172
Position from start codon	0.1475
Change in CAI (synonymous)	0.1975
Change in tAI (synonymous)	0.5445
Change in mRNA folding energy	0.0026

LITERATURE CITED

Agashe, D., N. C. Martinez-Gomez, D. A. Drummond, and C. J. Marx. 2013. Good codons, bad transcript: Large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. Mol. Biol. Evol. 30:549–560.

Agashe, D., M. Sane, K. Phalnikar, G. D. Diwan, A. Habibullah, N. C. Martinez-Gomez, V. Sahasrabuddhe, W. Polachek, J. Wang, L. M. Chubiz, and C. J. Marx. 2016. Large-effect beneficial synonymous mutations mediate rapid and parallel adaptation in a bacterium. Mol. Biol. Evol. 33:1542–1553.

Bailey, S. F., A. Hinz, and R. Kassen. 2014. Adaptive synonymous mutations in an experimentally evolved Pseudomonas fluorescens population. Nat. Commun. 5. Nature Publishing Group.

Bank, C., R. T. Hietpas, A. Wong, D. N. Bolon, and J. D. Jensen. 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.

Bataillon, T., T. Zhang, and R. Kassen. 2011. Cost of adaptation and fitness effects of beneficial mutations in pseudomonas fluorescens. Genetics 189:939–949.

Beisel, C. J., D. R. Rokyta, H. A. Wichman, and P. Joyce. 2007. Testing the extreme value domain of attraction for distributions of beneficial fitness effects. Genetics 176:2441–2449.

Bentele, K., P. Saffert, R. Rauscher, Z. Ignatova, and N. Blüthgen. 2013. Efficient translation initiation dictates codon usage at gene start. Mol. Syst. Biol. 9:675.

Betancourt, A. J. 2009. Genomewide patterns of substitution in adaptively evolving populations of the RNA bacteriophage MS2. Genetics 181:1535–1544.

Cambray, G., and D. Mazel. 2008. Synonymous genes explore different evolutionary landscapes. PLoS Genet. 4.

Chamary, J. V., and L. D. Hurst. 2005. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biol. 6:R75.

Chamary, J. V, J. L. Parmley, and L. D. Hurst. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat. Rev. Genet. 7:98–108.

Chaney, J. L., and P. L. Clark. 2015. Roles for synonymous codon usage in protein biogenesis. Annu. Rev. Biophys. 44:143–166.

Cortazzo, P., C. Cervenansky, M. Marin, C. Reiss, R. Ehrlich, and a Deana. 2002. Silent mutations affect in vivo protein folding in Escherichia coli. Biochem. Biophys. Res. Commun. 293:537–541.

Cuevas, J. M., P. Domingo-Calap, and R. Sanjuán. 2012. The fitness effects of synonymous mutations in DNA and RNA viruses. Mol. Biol. Evol. 29:17–20.

Deutscher, M. P. 2006. Degradation of RNA in bacteria: Comparison of mRNA and stable RNA. Nucleic Acids Res. 34:659–666.

dos Reis, M., R. Savva, and L. Wernisch. 2004. Solving the riddle of codon usage preferences: A test for translational selection. Nucleic Acids Res. 32:5036–5044.

Duret, L. 2000. tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. 16:287–289.

Engler, C., R. Kandzia, and S. Marillonnet. 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS One 3.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet. 8:610–8.

Firnberg, E., J. W. Labonte, J. J. Gray, and M. Ostermeier. 2014. A comprehensive, high-resolution map of a gene's fitness landscape. Mol. Biol. Evol. 31:1581–1592.

Fisher, R. A. 1930. The genetical theory of natural selection. Oxford University Press, Oxford.

Ford, M. J. 2002. Applications of selective neutrality tests to molecular ecology. Mol. Ecol. 11:1245–1262.

Fragata, I., S. Matuszewski, M. A. Schmitz, T. Bataillon, J. D. Jensen, and C. Bank. 2018. The fitness landscape of the codon space across environments. , doi: 10.1101/252395.

Frumkin, I., D. Schirman, A. Rotman, F. Li, L. Zahavi, E. Mordret, O. Asraf, S. Wu, S. Levy, and Y. Pilpel. 2017. Gene architectures that minimize cost of gene expression. Mol. Cell 65:In press. Elsevier Inc.

Gillespie, J. 1984. Molecular evolution over the mutational landscape. Evolution (N. Y). 38:1116–1129.

Gillespie, J. H. 1983. A simple stochastic gene substitution model. Theor. Popul. Biol. 23:202–215.

Gingold, H., and Y. Pilpel. 2011. Determinants of translation efficiency and accuracy. Mol. Syst. Biol. 7:481. Nature Publishing Group.

Good, B. H., M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai. 2017. The dynamics of molecular evolution over 60,000 generations. Nature 551:45–50. Nature Publishing Group.

Gouy, M., and C. Gautier. 1982. Codon usage in bacteria: Correlation with gene expressivity. Nucleic Acids Res. 10:7055–7074.

Gu, W., M. Li, Y. Xu, T. Wang, J.-H. Ko, and T. Zhou. 2014. The impact of RNA structure on coding sequence evolution in both bacteria and eukaryotes. BMC Evol. Biol. 14:87.

Hauber, D. J., D. W. Grogan, and R. W. DeBry. 2016. Mutations to less-preferred synonymous codons in a highly expressed gene of Escherichia coli: Fitness and epistatic interactions. PLoS One 11:1–16.

Hershberg, R., and D. A. Petrov. 2008. Selection on codon bias. Annu. Rev. Genet. 42:287–299.

Hietpas, R. T., J. D. Jensen, and D. N. a Bolon. 2011. Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. U. S. A. 108:7896–7901.

Hilterbrand, A., J. Saelens, and C. Putonti. 2012. CBDB: the codon bias database. BMC Bioinformatics 13:62. BioMed Central Ltd.

Hughes, A. L., and M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170.

Hunt, R. C., V. L. Simhadri, M. Iandoli, Z. E. Sauna, and C. Kimchi-Sarfaty. 2014. Exposing synonymous mutations. Trends Genet. 30:308–321. Elsevier Ltd.

Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2:13–34.

Imhof, M., and C. Schlotterer. 2001. Fitness effects of advantageous mutations in evolving Escherichia coli populations. Proc. Natl. Acad. Sci. 98:1113–1117.

Joyce, P., D. R. Rokyta, C. J. Beisel, and H. A. Orr. 2008. A general extreme value theory model for the adaptation of DNA sequences under strong selection and weak mutation. Genetics 180:1627–1643.

Kanaya, S., Y. Yamada, M. Kinouchi, Y. Kudo, and T. Ikemura. 2001. Codon usage and tRNA genes in eukaryotes: Correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. J. Mol. Evol. 53:290–298.

Kassen, R., and T. Bataillon. 2006. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat. Genet. 38:484–488.

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

Kimchi-Sarfaty, C., J. M. Oh, I.-W. Kim, Z. Sauna, A. M. Calcagno, S. Ambudkar, and M. Gottesman. 2007. A "Silent" Polymorphism in the MDR1 Gene Changes Substrate Specificity. Science (80-.). 315:525–529.

Knöppel, A., J. Näsvall, and D. I. Andersson. 2016. Compensating the fitness costs of synonymous mutations. Mol. Biol. Evol. 33:1461–1477.

Komar, A. A., T. Lesnik, and C. Reiss. 1999. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation. FEBS Lett. 462:387–391.

Kryazhimskiy, S., and J. B. Plotkin. 2008. The population genetics of dN/dS. PLoS Genet. 4.

Kudla, G., A. W. Murray, D. Tollervey, and J. B. Plotkin. 2009. Coding-sequence determinants of gene expression in Escherichia coli. Science (80-.). 324:255–258.

Lawrie, D. S., P. W. Messer, R. Hershberg, and D. A. Petrov. 2013. Strong purifying selection at synonymous sites in D. melanogaster. PLoS Genet. 9:33–40.

Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler. 1991. Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. Am. Nat. 138:1315–1341.

Li, C., W. Qian, C. J. Maclean, and J. Zhang. 2016. The fitness landscape of a tRNA gene. Science (80-.). 352:837–840.

Li, G.-W., E. Oh, and J. S. Weissman. 2012. The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature 484:538–541.

Lind, P. A., and D. I. Andersson. 2013. Fitness costs of synonymous mutations in the rpsT gene can be compensated by restoring mRNA base pairing. PLoS One 8:1–6.

Lind, P. A., O. G. Berg, and D. I. Andersson. 2010. Mutational robustness of ribosomal protein genes. Science (80-.). 330:825–827.

Marin, M. 2008. Folding at the rhythm of the rare codon beat. Biotechnol. J. 3:1047–1057.

Martin, G., and T. Lenormand. 2008. The distribution of beneficial and fixed mutation fitness effects close to an optimum. Genetics 179:907–916.

Maynard Smith, J. 1970. Natural selection and the concept of a protein space. Nature 225:563–564.

McDonald, M. J., T. F. Cooper, H. J. E. Beaumont, and P. B. Rainey. 2011. The distribution of

fitness effects of new beneficial mutations in Pseudomonas fluorescens. Biol. Lett. 7:98–100.

Mohammad, F., C. J. Woolstenhulme, R. Green, and A. R. Buskirk. 2016. Clarifying the translational pausing landscape in bacteria by ribosome profiling. Cell Rep 14:686–694.

Navon, S., and Y. Pilpel. 2011. The role of codon selection in regulation of translation efficiency deduced from synthetic libraries. Genome Biol. 12:R12. BioMed Central Ltd.

Nielsen, R. 2005. Molecular Signatures of Natural Selection. Annu. Rev. Genet. 39:197–218.

Orr, H. A. 2003. The distribution of fitness effects among beneficial mutations. Genetics 163:1519–26.

Orr, H. A. 2006. The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. J. Theor. Biol. 238:279–285.

Orr, H. A. 2005. The genetic theory of adaptation: A brief history. Nat. Rev. Genet. 6:119–127.

Orr, H. A. 2002. The population genetics of adaptation: The adaptation of DNA sequences. Evolution (N. Y). 56:1317–1330.

Plotkin, J. B., and G. Kudla. 2011. Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 12:32–42. Nature Publishing Group.

Presnyak, V., N. Alhusaini, Y. H. Chen, S. Martin, N. Morris, N. Kline, S. Olson, D. Weinberg, K. E. Baker, B. R. Graveley, and J. Coller. 2015. Codon optimality is a major determinant of mRNA stability. Cell 160:1111–1124. Elsevier Inc.

Rokyta, D. R., C. J. Beisel, P. Joyce, M. T. Ferris, C. L. Burch, and H. A. Wichman. 2008. Beneficial fitness effects are not exponential for two viruses. J. Mol. Evol. 67:368–376.

Rokyta, D. R., P. Joyce, S. B. Caudle, and H. A. Wichman. 2005. An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat. Genet. 37:441–444.

Rozen, D. E., J. A. G. M. De Visser, and P. J. Gerrish. 2002. Fitness effects of fixed beneficial mutations in microbial populations. Curr. Biol. 12:1040–1045.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sanjuán, R., A. Moya, and S. F. Elena. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. U. S. A. 101:8396–401.

Saunders, R., and C. M. Deane. 2010. Synonymous codon usage influences the local protein

structure observed. Nucleic Acids Res. 38:6719–6728.

Schoustra, S. E., T. Bataillon, D. R. Gifford, and R. Kassen. 2009. The properties of adaptive walks in evolving populations of fungus. PLoS Biol. 7.

Sharp, P. M., E. Cowe, D. G. Higgins, D. C. Shields, K. H. Wolfe, and F. Wright. 1988. Codon usage patterns in Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster and Homo sapiens; a review of the considerable within-species diversity. Nucleic Acids Res. 16:8207–8211.

Sharp, P. M., and W.-H. Li. 1987. The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15:1281–1295.

She, R., and D. F. Jarosz. 2018. Mapping causal variants with single-nucleotide resolution reveals biochemical drivers of phenotypic change. Cell 172:478–490.e15. Elsevier Inc.

Sorensen, M. A., C. G. Kurland, and S. Pedersen. 1989. Codon usage determines translation rate in Escherichia coli. J. Mol. Biol. 207:365–377.

Stahl, E. A., and J. G. Bishop. 2000. Plant–pathogen arms races at the molecular level. Curr. Opin. Plant Biol. 3:299–304.

Still, A. W., and A. P. White. 1981. The approximate randomization test as an alternative to the F test in analysis of variance. Br. J. Math. Stat. Psychol. 34:243–252.

Takanami, M., and G. Zubay. 1964. An Estimate of the Size of the Ribosomal Site for Messenger RNA Binding Author (s): M . Takanami and G . Zubay Source : Proceedings of the National Academy of Sciences of the United States of America , Published by : National Academy of Sciences Stable. Proc. Natl. Acad. Sci. U. S. A. 51:834–839.

Tsai, C., Z. E. Sauna, C. Kimchi-Sarfaty, S. V Ambudkar, M. Gottesman, and R. Nussinov. 2008. Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. J. Mol. Biol. 383:281–291.

Tuller, T., A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman, and Y. Pilpel. 2010. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141:344–354. Elsevier Ltd.

Wiser, M. J., and R. E. Lenski. 2015. A comparison of methods to measure fitness in Escherichia coli. PLoS One 10:1–11.

Xiang, Y., S. Gubian, S. Suomela, and J. Hoeng. 2013. Generalized simulated annealing for

efficient global optimization: the GenSA package for R. R J. 5.

Yu, C.-H., Y. Dang, Z. Zhou, C. Wu, F. Zhao, M. S. Sachs, and Y. Liu. 2015. Codon usage influences the local rate of translation elongation to regulate co-translational protein folding. Mol. Cell 59:744–754.

Zanini, F., V. Puller, J. Brodin, J. Albert, and R. A. Neher. 2017. In vivo mutation rates and the landscape of fitness costs of HIV-1. Virus Evol. 3.

Zhou, Z., Y. Dang, M. Zhou, L. Li, C.-H. Yu, J. Fu, S. Chen, and Y. Liu. 2016. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. Proc. Natl. Acad. Sci. U. S. A. 113:E6117–E6125.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406–3415.