# The dynamics of adaptive evolution in fluctuating environments

by

Farah Abdul-Rahman

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Biology

New York University

May 2021

_____

David Gresham, PhD

# DEDICATION

...To my family...

# ACKNOWLEDGEMENTS

Learning is the biggest gift a person can receive and so I am thankful to everyone I have crossed paths with during my time at NYU and the lasting effect they have had on me. I thank my advisor David Gresham for his mentorship and for having me in his lab *where* I learned about the awesome power of yeast genetics. I also thank Daniel Tranchina for closely collaborating with me and under his mentorship I learned more about statistics and mathematical modeling than I could have hoped. I thank all my committee members who have each been pivotal in guiding my project and progress, Gloria Corruzzi, Christine Vogel, Duncan Smith, and Jef Boeke. I thank all past and present lab members who have been instrumental in building the everyday learning culture in the lab including, in no particular order, Nathan Brandt, Benjamin Neymotin, Pieter Spealman, Angela Hickey, Chris Jackson, Stephanie Lauer, Darach Miller, Rodoniki Athanasiadou, Siyu Sun, Grace Avecilla, and several others whom I have briefly overlapped with. I thank all the amazing mentees I have had the opportunity to work with, to teach and to learn from including Erika Levine, Janell Womack, Carla Rodriguez, and Charles Miller. I thank Justin Rendleman, Dayanne Castro, Gorkem Garipler and Tobi Schraink for their titillating scientific conversations and their friendship and support. I thank the rest of the biology department at NYU including graduate students, postdocs, and faculty and in particular, Justin Blau, and Steve Small who have been the soul of the department. I thank Mark Siegal whom I TA-ed for for the majority of my PhD and learned so much about teaching from. I also thank other faculty for discussions and conversations including Sevinc Ercan. I thank GenCore past and

present members including Tara Reid and Olivia Micci-Smith for their leadership. I especially thank Hana Husic for her friendliness and commitment to helping others from the fine technical details to experimental design. I thank the administrative staff including Sherry Scott, Lita Robinson, Anne Brown and many others who were integral in smoothing out the kinks whenever they arose, and consistently led with kindness and support.

# ABSTRACT

Microbes live in dynamic environments that present ongoing challenges to reproduction and survival. One class of environmental change common in the natural world is periodic fluctuations such as seasonal changes in temperature, precipitation, and nutrient availability. Unlike static conditions in which a selective pressure is continuously experienced by the organism, periodically fluctuating conditions cycle between different selective pressures resulting in complex evolutionary dynamics. Whereas evolutionary dynamics in static environments have been extensively characterized using experimental evolution, the empirical study of how fluctuating environments impact the dynamics and outcomes of evolution remains largely unexplored. In this thesis, I address the questions of how fluctuating environments affect population diversity, fitness dynamics, and mutation supply using the budding yeast *Saccharomyces cerevisiae*. In chapter 1, I review literature detailing how fluctuating environments influence adaptive evolution in microbes. In chapter 2, I test the effect of fluctuations in nutrient concentration with different periods using a synthetic population comprising thousands of different genotypes to show that environmental fluctuations result in the maintenance of increased genetic diversity in comparison to static conditions. I demonstrate that this feature is driven by fluctuating conditions enriching for genotypes with neutral fitness and a class of genotypes that follows a sinusoidal growth dynamics. In static conditions, diversity is reduced as a result of strong selective sweeps in which a few highly fit genotypes rapidly rise to high frequency. In chapter 3, I used experimental evolution of clonal populations in which *de*

*novo* mutation is the sole source of heritable variation, to study how environments that fluctuate in nutritional quality affect the dynamics of copy number variants (CNVs). I describe the identification of a subpopulation that acquires CNVs at multiple loci uniquely in fluctuating environments. In chapter 4, I present the development of a new lineage tracking molecular barcoding scheme that imposes a reduced genetic load on cells compared with existing DNA lineage tracking systems. Together, these studies underscore the importance of understanding how periodically fluctuating environments determine evolutionary outcomes and the principles that govern these effects. The implications of these findings are summarized and discussed in Chapter 5.

# TABLE OF CONTENTS


Dedication i
Acknowledgements ii
Abstract iv
List of Figures x
List of Tables xi
Chapter 1: Introduction 1
1.1: Abstract 1
1.2: Fluctuating selection is universal in the natural world 1
1.3: Characteristics of temporally fluctuating environments 2
1.3.1: Predictability 2
1.3.2: Frequency 4
1.3.3: Magnitude 5
1.3.4 Interaction of simultaneous fluctuations 6
1.4: Adaptive strategies in periodically fluctuating environments 5
1.4.1: Sense-and-response 7
1.4.2: Memory 9
1.4.3: Biological clocks 10
1.4.4: Diversification 11
1.5: Theoretical aspects of fluctuating selection 15
1.6: Conclusion 18

Chapter 2: Fluctuating environments maintain genetic diversity through neutral fitness effects and balancing selection 20
2.1: Abstract 20
2.2: Introduction 21
2.3: Results 25
2.3.1 Modeling nutritional fluctuations in chemostats 27
2.3.2 Selection on heterogeneous populations in a chemostat 30
2.3.3 Fluctuating environments maintain greater genetic diversity 31
2.3.4 Genotypes exhibit distinct frequency dynamics 38
2.3.5 Environmental fluctuations select for specific mutant classes 41
2.3.6 Switching conditions harbor the highest short-term growth rates 45
2.4: Discussion 46
2.5: Methods 48
2.5.1 Media 48
2.5.2 Experimental measurements of model parameters 48

2.5.3 Mathematical modeling of chemostat growth in fluctuating environments 49
2.5.4 Culturing conditions 49
2.5.5 Analysis of Barseq data 49
2.5.6 Mathematical modeling of genotype behavior 50
2.6: Supplementary figures 51
Chapter 3: The dynamics of copy number variants in fluctuating environments 64
3.1: Abstract 64
3.2: Introduction 65
3.3: Results 68
3.3.1 Fluorescence is a proxy measurement for gene copy number at two distinct loci 70
3.3.2 Periodic fluctuations in nitrogen quality do not alter cell size and density 71
3.3.3 CNV evolutionary dynamics are condition- and genetic locus-dependent 74
3.3.4 Different gene copy numbers per cell are the drivers of adaptation at different temporal phases 76
3.3.5 A dual-color CNV reporter increases precision in detecting true CNV events 77
3.3.6 Fluctuating environments select for CNV-generalists 78
3.4 Future and ongoing work 80
3.5: Discussion 83
3.6: Methods 84
3.6.1 Media and growth conditions 84
3.6.2 Strains construction 85
3.7: Supplementary figures 87
Chapter 4: Constructing a new lineage tracking barcode system in budding yeast 91
4.1: Abstract 91
4.2: Introduction 91
4.3: Results 96
4.3.1 Overview of new barcode lineage tracking system 96
4.3.2 Constructing a Gal-inducible Cas9 yeast strain 98
4.3.3 Constructing gRNA plasmids 100
4.4 Future work: Creation of a high-complexity barcode plasmid library 101
4.5: Discussion 103
4.6: Conclusion 104
4.7: Methods 105

4.7.1 Construction of plasmids 105
Chapter 5: Conclusion 107
5.1: Summary 107
5.2: Future directions 110
References 113
Appendix: Supplementary methods for Chapter 2 128

# LIST OF FIGURES

Figure 2.3.1 Periodically fluctuating environments in the chemostat 26
Figure 2.3.2 Fluctuating selective conditions maintain greater genetic diversity than static selective conditions 29
Figure 2.3.3 Diverse temporal trajectories of genotypes in different selective conditions 37
Figure 2.3.4. A subset of genotypes have a predictive relationship between fluctuating and static selective conditions 43
Figure 2.3.5. The switch condition uniquely results in short term fitness changes that are not detected over larger timescales 45
Figure 2.6.1 Modelling the growth of four thousand genotypes in the chemostat 52
Figure 2.6.2 Barseq library quality control 53
Figure 2.6.3 Maintenance of genetic diversity in different fluctuating selections 54
Figure 2.6.4 Library size does not affect diversity estimates 55
Figure 2.6.5 The rate of change in total strain number across all conditions 56
Figure 2.6.6 Strain extinction profiles in each condition 57
Figure 2.6.7 Change in percent population proportion for pulse conditions 58
Figure 2.6.8 Diversity measurements of experiments from the Salingnon dataset 59
Figure 2.6.9 Genotype dynamics in static and fluctuating environments 60
Figure 2.6.10 Gene set enrichment analysis (GSEA) of fitness effects in each Condition 61
Figure 2.6.11 Deletion of the ERAD genes uniquely results in increased fitness in fluctuating environments 62

Figure 3.3.1 The establishment of a model system for studying adaptive evolution of CNVs in fluctuating selection 70
Figure 3.3.2 A dual-fluorescence CNV-reporter system simultaneously tracks evolutionary dynamics of CNVs at two loci in static and fluctuating conditions 73
Figure 3.3.3 Muller plots of copy number variants detected using the GAP1-PUT4 CNV reporter 76
Figure 3.3.4 The proportion of CNV specialists, and generalists in the GAP1-PUT4 CNV reporter 79
Figure 3.7.1 Control strains experimentally evolved in three distinct conditions for 250 generations 88
Figure 3.7.2 Forward scatterm a proxy measurement for cell size, shows minimal change across conditions 89
Figure 3.7.3 Muller plots of copy number mutations in the GAP1-MEP2 CNV reporter 90
Figure 3.7.4 Plots showing proportion of CNV specialists, and generalists in the GAP1-MEP2 CNV reporter 91

Figure 4.2.1. Levy et al's barcoding system 94
Figure 4.2.2. Fitness assays of two genotypes barcoded with the Levy barcoding system 96
Figure 4.3.1. Lineage barcoding design overview 99
Figure 4.3.2. Insertion of galactose-inducible Cas9 at a neutral locus 101
Figure 4.3.3. Golden gate performed to create the gRNA plasmid 102
Figure 4.4.1. Completion of the barcode system 103

# LIST OF TABLES

Table 2.6.1 Pairwise correlation matrix of counts across all conditions 62
Table 2.6.2 DFE statistical measurements 62
Table 4.6.1 Plasmids used in this chapter 103

I have worked on various additional projects during my graduate studies, a number of which have resulted in publications. I have contributed to a project that aimed to understand how motor neurons are programmed from embryonic stem cells by performing single-cell RNAseq (Velasco et al. 2017). I contributed to a study aimed at understanding the transcriptional response in *S. cerevisiae* to transient changes in nitrogen concentration by performing RNAseq analysis of cells in nitrogen-rich and nitrogen-poor conditions (Airoldi et al. 2016). I assisted in characterizing isolates from an adaptive evolution experiment by performing qPCR, to quantify allele copy number, and by performing fitness assays (Hong et al. 2018). I contributed to a protocol for the measurement of RNA degradation rates (Abdul-Rahman and Gresham 2016). These studies are not contained within this thesis.

# Chapter 1: Introduction

## 1.1 Abstract

There is no environment that is truly static and as such microbes have adapted to respond to fluctuations through a number of mechanisms that maximize their fitness. Fluctuations are defined by several characteristics such as their predicitiblity, frequency, and magnitude, that determine the strategy that is best suited as a response. Here, I review the existing literature on fluctuating selection with a focus on microbial populations and budding yeast and with the goal of identifying similarities and differences between varieties of fluctuating selection in the context of evolution. I focus on the effects of fluctuating environments on genetic diversity from the perspective of experimental studies and the theoretical evolutionary framework.

## 1.2 Fluctuating selection is universal in the natural world

Fluctuating selection is of particular interest in the field of evolution as no environment is truly static and therefore it is impossible to fully understand the processes and outcomes of evolution without considering temporal changes in selection. Because of the complexities posed by introducing changes across the dimension of time, evolution studies have frequently focused on static environments. The term fluctuating selection encompasses a wide variety of scenarios, which raises the question of whether having an all encompassing term to describe these diverse scenarios is useful and if constructing a unified framework is possible.

Fluctuating selection in its broadest definition encompasses any changes in either biotic or non-biotic factors that present selective pressures on organismal survival and reproduction. Non-biotic factors include temperature, chemical concentrations (e.g. salt), and pH, whereas biotic factors include population density, predation, and sexual mate availability (Morand and Lajaunie 2018). Environmental variability can either be spatial, such as in the case of chemical gradients that are frequently found in natural environments such as soil sediments, or can be temporal, such as with tidal regions on coastlines (Oliveira de Santana et al. 2021). Spatial heterogeneity frequently results in niche specialization in populations, which effectively allows different genotypes to consistently only experience one dimension of the environment (Levins 1968). Here, I focus on the impacts of temporal fluctuations on microbial populations to understand defining characteristics and their effects on the process of selection and outcomes of evolution.

## 1.3 Characteristics of temporally fluctuating environments

1.3.1: Predictability: Foreseeing an upcoming environmental change is beneficial during cellular decision-making in order to maximize individual fitness. Cells that can anticipate upcoming stimuli earlier than other cells, have time to ‘prepare’, which can confer an advantage over ‘unprepared’ cells. One example of anticipatory preparation in budding yeast was observed in natural isolates grown in a mixture of two different carbon sources, glucose and galactose. Yeast consume glucose preferentially and undergo a lag phase before switching to growth on the less preferred carbon source, a phenomenon known as the crabtree effect. In this study it was found that despite

consuming glucose first, cells began producing galactose utilization genes hours before glucose was depleted, effectively rewiring their machinery to rapidly switch to the utilization of the alternate carbon source upon glucose depletion. This study also showed that in diverse natural isolates, this response was variable indicating that such responses are heritable and vary within species. This suggests that anticipation is not always the preferred and that different environments may select for different strategies, or that having a variety of strategies is in itself useful for a species (J. Wang et al. 2015).

Predictability of environmental change is a consequence of either 1) a regularly repeating pattern of change, or 2) sensing a stimulus that precedes, but is coupled to that change. A striking example of this later mechanism is the transcriptional coupling of the heat-shock response regulon to the aerobic functions in *Escherichia coli*. The rationale follows that when *E. coli* cells are outside a host's body, the temperature is low, and oxygen is high, but when they enter the animal host's mouth the temperature is higher. As *E. coli* cells go further into the GI tract towards the gut, oxygen levels decrease. Therefore coupling the transcription of the temperature response and oxygen-availability response permits cells to only need to sense a signal - change in temperature - to modulate a response for an impending change in oxygen availability. This anticipatory shortcut bypasses the need for sensory information to be relayed into the cell twice for an appropriate response and eliminates a potential lag phase that may be associated with cellular remodeling that occurs as a direct response to changes in oxygen (Tagkopoulos, Liu, and Tavazoie 2008).

On the other hand, random fluctuations frequently require some probabilistic behavioral response such as the diversification of strategies, which generally achieves lower fitness gains for a population. In the absence of predictive information, cells must blindly diversify their response instead, which means that several strategies that individuals will try are destined to fail resulting in a guaranteed cost. The ability to predict change allows cells to focus their resources on a single strategy which severely decreases any associated cost if the prediction is correct.

1.3.2: Frequency: Another important characteristic of change is how often it occurs, which can have profound effects on an organism's fitness and optimal strategic response (Walworth et al. 2020). The most obvious way in which frequency is important, is that the most recurrent fluctuations in nature are likely to be experienced more by cells meaning that they will have a greater fitness effect whether the fluctuation is stressful, and therefore costly, or a surplus of a resource and therefore beneficial. For example, microbes exposed to harsh conditions frequently encounter changes in temperature, oxidative stress, and heavy metals, all of which increase protein misfolding rendering them nonfunctional or malfunctioning requiring a coping mechanism. The heat-shock response, which is conserved across all domains of life, increases the number of molecular chaperones to alleviate the negative effects of stressors on proteins. On the other hand, cells may adapt to make use of regularly fluctuating resources such as in the case of cyanobacteria's dependence on sunlight and are tuned to the diurnal light cycle to photosynthesize and fix carbon. However, it has also been shown that the frequency of fluctuation is important for the dynamics in which evolution

proceeds as well as for its outcomes. In *E. coli*, populations exposed to either rapid or sparse antibiotic pulses are more likely to develop resistance through selective sweeps in comparison to pulses happening at intermediate length intervals (Lin and Kussell 2016). Frequency alone cannot determine the strength of selection as the magnitude of the event can affect fitness differentially.

1.3.3: Magnitude: The magnitude of environmental change relates to whether a change occurs in a mild or severe burst. Magnitude pertains to variables that are continuous, such as nutrient concentration, as opposed to discrete, such as nutrient quality. For example, a recurrent burst in nutrient concentration, could significantly relax selective pressure for cells in comparison to recurrent mild changes in concentration which may have a minimal effect on fitness. In contrast, if a nutrient source becomes depleted, which could lead to starvation and death, it imposes a large fitness effect, and therefore adapting a response mechanism regardless of frequency of occurrence is essential. Indeed microbes have ubiquitously evolved numerous adaptive strategies to mitigate the effect of starvation and to survive cycles of depletion and surplus of nutrients. It has been suggested that using the response parameters of an organism to a stimulus allows for back-tracing and the determination of the environmental characteristics in which the organism has evolved (Tagkopoulos, Liu, and Tavazoie 2008). Sporulation and quiescence are two widely used strategies across the microbial kingdom to cope with nutrient depletion. While these are extremes in magnitude of a fluctuation, realistically, they likely occur less frequently than mild to moderate changes. As such, microbes have evolved a tight gene network of interaction to modulate the

nutrient-availability to growth-response to respond to mild-moderate fluctuations in concentration.

1.3.4 Interaction of simultaneous fluctuations: Each different characteristic of fluctuating environments independently poses significant challenges for a cell; however, in natural environments cells face fluctuations comprising a mixture of these characteristics such that a fluctuation may be recurrent at regular intervals but vary in magnitude. Yeast in natural environments for example, have to manage changes in temperature that cycle between night and day while simultaneously having to manage changes in nutrient availability as well as pH level, each of which fluctuates at a different rate, magnitude and level of predictability. This poses a significant challenge in decision-making in resource allocation for cells especially when energy is a limited resource. Modular gene expression networks are an important system to integrate different types of simultaneous fluctuations, and are likely an adaptation to allow rapid coupling and decoupling of modules to respond to environmental fluctuations. The study of how microbes respond, adapt, and evolve in fluctuating environments remains in its infancy and extensive research is required to understand the underlying principles.

## 1.4 Adaptive strategies in periodically fluctuating environments

Environmental fluctuations pose an enormous challenge for the survival of microorganisms since the majority are single-celled and are non-motile and directly exposed to the surrounding environment. As such several strategies have been evolved to deal with a variety of fluctuations that fall under one of four classes.

1.4.1: Sense-and-response: Sense-and-response refers to strategies in which cells actively monitor their external environment and relay information to alter gene expression and cellular functions. Sense-and-response strategies don't require previous expectations of how the environment will change allowing them to be broadly useful for a variety of factors. This strategy, however, comes at the price of energy being used for constitutive sensing and it also results in a lag in response that is associated with the time needed to relay sensory information and appropriately remodel the cell.

Many mechanisms have been evolved to sense the environment. In bacteria, the two-component regulatory system which is composed of a membrane-bound histidine kinase that senses the environment and a response regulator that mediates the cellular response. *S. cerevisiae* has nutrient-sensing systems that rely on transceptors, membrane-bound receptors that are homologous to nutrient transporters some of which maintain their transporting functionality while it is lost in others. Four main transceptors are responsible for nutrient-sensing in budding yeast, Ssy1, Snf3, Rgt2, and Mep2 (Lin and Kussell 2016; Chantranupong, Wolfson, and Sabatini 2015). Snf3 senses low levels of glucose (Bisson et al. 1987) and Rgt2 senses high levels of glucose (Bisson et al. 1987; Ozcan et al. 1996). Ssy1 detects amino acids and activates a signalling pathway which upregulates amino acid metabolism genes. MEP2 is known for its role in sensing and importing ammonium into the cell. Additionally, it controls pseudohyphal growth in low-ammonium conditions enabling nutrient scavenging (Lorenz and Heitman 1998).

In addition, the cell also has internal sensors. For example, in the absence of internal glucose, the SNF1 complex is activated to regulate the expression of cellular machinery responsible for the carbon stress response such as metabolic enzymes involved in fatty acid metabolism and carbohydrate storage (Walworth et al. 2020; Persson et al. 2020). In a similar vein, the target of rapamycin complex 1 (TORC1) is a regulator of growth in response to nitrogen availability. TORC1 is linked to a variety of cellular mechanisms that are essential for growth, including ribosomal proteins. This high degree of connectivity is useful in coordinating different processes in response to nutrient availability, however, this also results in significant costs to the cell highlighting some of the disadvantages of sense-and-response strategies.

The Environmental Stress Response (ESR) is another example of how the environment can trigger responses that result in extensive transcriptional and translational remodelling of the cells  requiring the re-allocation of energy and resources. Another way in which sense-and-response strategies can impose a cost is because they usually have a lag phase, the time from initial environmental sensing to the completion of cellular response. Lag phases are well studied in microbes such as with the lac operon in which *E. coli* grown on glucose and then switched to lactose have a phase of slowed growth, or such as with the Crabtree effect in which yeast is known to consume glucose preferentially and undergo a lag phase before switching to growth on ethanol. This has also been observed in response to changes in environmental nitrogen in which yeast growing on a poor nitrogen source and then switched to a rich nitrogen source undergo a period of a slowed growth in which extensive transcriptional

remodelling takes place, including changes in RNA degradation rates of nutrient transporters such as the amino acid transporter, GAP1, and the proline transporter, PUT4 (D. Miller, Brandt, and Gresham 2018; Airoldi et al. 2016).

1.4.2: Memory: The idea that an organism can learn from the previous environment and use that knowledge to prepare for an upcoming change is important in the context of changing environments. Memory responses and their benefits have been observed in multiple microorganisms. In *E. coli*, cells that are grown in glucose and then switch to growth on lactose exhibit a lag phase. However, if these same cells are exposed to glucose and then lactose again, the lag phase is shortened in comparison to cells that have never been exposed to glucose before (Lambert and Kussell 2014). This response shows that cells that have been exposed to an environment before have kept a memory so that now they are better equipped to deal with similar conditions in the future. This response is mediated at the molecular level by a slowed degradation rate of the protein lacY which is one of the core components of the lac operon that is responsible for lactose metabolism in *E. coli*. This system is a great model for studying how molecular memory works in microorganisms to increase fitness and support for this model has been identified in yeast as well (Razinkov et al. 2013). Similarly in yeast, galactose elicits a memory response as well in individuals that have been previously grown in galactose and then switch to glucose and then galactose again. Additionally, a more nuanced type of memory is one in which being exposed to a mild stress allows protection from a more lethal stress in the future (Berry and Gasch 2008).

1.4.3 Biological clocks: The most well-known biological clock is the circadian clock which regulates the internal cellular environment with the day and night light cycle. Circadian rhythms have been observed in animals, plants, insects, and microbes. Cyanobacteria are a class of photosynthesizing bacteria that benefit from having a circadian rhythm because they rely on the presence of light to generate energy. Recent evidence also suggests that the gut microbiome helps regulate the intestinal circadian rhythm (Mukherji et al. 2013) and that the relative abundance and composition of the gut microbiome exhibit diurnal oscillations which in turn help maintain the host's gut circadian rhythm (Thaiss et al. 2014, 2016).

True circadian rhythms in budding yeast have been elusive, however, there is evidence for clock-like oscillatory behavior which occurs over timescales smaller than 24 hours, termed ultradian rhythms (Mukherji et al. 2013; Robertson et al. 2008; Tu et al. 2005). Additionally, budding yeast have been shown to be entrained by 24-hour temperature oscillations. After such a regime, RNA expression of the two permeases, MEP2 and GAP1 oscillated at a similar interval to the entrainment length (Eelderink-Chen et al. 2010). Evolving biological clocks seems to be the optimal solution to environmental fluctuations that occur at regular intervals and are very predictable as they improve upon the sense-and-response strategies by not requiring a constitutive sensor and therefore using less energy. Clocks however are very susceptible to situations in which environments do not reliably oscillate at repetitive intervals, such scenarios would more closely resemble random fluctuations.

1.4.4: Diversification: Fluctuations that are more challenging to predict require a different cellular approach to maximize fitness. A common approach is for individuals in a population to use an assortment of strategies to increase the probability that one would be suitable to increase fitness. This can manifest by individuals either randomly switching between different responses following some probabilistic value, or through different individuals consistently displaying diverse responses. This adaptation can either manifest through phenotypic diversification, in which cells are isogenic but expression is stochastic, or through genetic diversity, in which a variety of genotypes that express different phenotypes are maintained within a population.

Phenotypic diversification: At the individual level, noisy gene expression is an important way in which variation is achieved. In *S. cerevisiae*, it has been observed that many genes with TATA boxes have noisy expression, and these genes are associated with the stress response (Gasch et al. 2000). Mutations in TATA boxes that decrease the noisiness of gene expression also decrease fitness in stressful conditions highlighting the importance of variation in responding to environmental change (Gasch et al. 2000; Blake et al. 2006). Galactose utilization genes are another example in which the level of noise is an important mode of regulation. When cells are grown in galactose alone, Gal genes show less noisy expression, however, growth in mixtures of galactose and glucose results in their bimodal expression which further supports that tuning the noise of gene expression is important in responding to environmental nutrient content (Healey, Axelrod, and Gore 2016) .

In isogenic populations, having a variety of phenotypes increases the robustness of the population in the face of environmental change which results in an increase in fitness of all the individuals. Bet-hedging has been observed in many instances in microbes. Stochastic switching in gene expression ensures that within clonal populations there remains subpopulations that display altered behaviors. One example is that of persister cells in genetically identical *E. coli* that encounter antibiotics. In nutrient rich environments a fraction of cells have been observed to grow at a slower rate than the average. These subpopulations, however, are able to withstand harsh environments in which antibiotics are administered that kill faster growing cells (Bigger 1944; Veening, Smits, and Kuipers 2008). A similar trend has been identified in *S. cerevisiae* in which a fraction of slow growing cells in nutrient-rich conditions can withstand heat shock and continue to proliferate despite the death of other cells that can ordinarily grow rapidly in nutrient rich conditions (Levy, Ziv, and Siegal 2012). These examples show how genetically identical populations can rely on phenotypic variation to survive in response to environmental stressors.

Genetic diversification: Populations also maintain a basal level of genetic variants, a phenomenon driven by the combination of new mutations and the strength of selection imposed by external forces. During experimental evolution it has been observed that several microbial species evolve a hypermutator genotype, an individual in which a mutation results in an increased error rate during replication. During the long-term evolution Lenski experiments, certain lineages of *E. coli* adapted by selecting for hypermutators (Sniegowski, Gerrish, and Lenski 1997) in which the mutation rate is

under selection and methyl-mismatch repair genes such as *mutS* are inactivated resulting in an increased error rate per replication cycle (Rosche and Foster 1999; Kang et al. 2019). This has also been observed during serial passaging of *Cryptococcus neoformans*, a fungal species that is a human pathogen. A D270G amino acid substitution in the exonuclease proofreading domain of DNA polymerase POL3 n (Magditch et al. 2012; Boyce et al. 2020) resulted in an increased mutation rate and increased likelihood of phenotypic switching which enabled spontaneous resistance to the chemical FK506 (Magditch et al. 2012). Increasing the error rate of the polymerase to increase the genetic diversity of a population has also been observed in poliovirus and further suggests that such an adaptation is important in adaptation to environmental change. Despite this, hypermutators have not been observed to arise and reach measurable frequencies during experimental evolution in *S. cerevisiae* although this is likely due to the observation that haploid hypermutators do not have a fitness advantage over nonmutator genotypes (Thompson, Desai, and Murray 2006).

While hypermutators generally increase mutation rate across the entire genome without locus-specific bias, increased mutation rates at a genetic locus with potential large-fitness effects may be beneficial. Mutation hotspots have been identified that result from specific genetic architectures, such as tandem repeats, Z DNA (McKinney et al. 2020), tRNA sequences, and long-terminal repeats. Copy number variants (CNVs), amplifications or deletions of a DNA locus, are frequently flanked by these architectures. The ribosomal DNA locus which is composed of a repeating array of 100-200 copies of a ~9.1Kb repeat unit, can expand and shrink during replication cycles in response to

environmental signals through the co-opting of extrachromosomal circles (Mansisidor et al. 2018) suggesting that CNVs are readily employed for rapid intergenerational adaptation. The copper-inducible CUP1 array of 2-20 copies of a ~2Kb repeat unit is another case in which the connection between environment signalling and number of gene copies has been established in which tandem repeats can expand or shrink in response to copper availability (Salim et al. 2021).

CNV generation in continuous culture using chemostats has been extensively studied using nutrient transporters as models of evolution. Under static nutrient-limiting conditions the transporters GAP1, MEP2, PUT4, SUL1 and DUR3 have been shown to undergo amplification when haploid *S. cerevisiae* is grown in conditions limited for the respective nutrient that they import (Hong and Gresham 2014; Payen et al. 2014; Gresham et al. 2008). These studies have focused on understanding how CNVs are generated and have discovered novel mechanisms such as Origin-Dependent Inverted-Repeat Amplification (ODIRA) as well as implicating previously known mechanisms such as non allelic homologous recombination and chromosomal translocations. Another goal has been to understand the dynamics in which they are generated and selected and their architectural diversity. It has been observed that during experimental evolution in static conditions, CNV generation and selection occurs in a highly repeatable manner and are produced through a variety of mechanisms in the population (Lauer et al. 2018; Spealman, Burrell, and Gresham 2020). Because of the large fitness effect of CNVs, their rapid generation, and their reversible nature, further

studies to understand the dynamics of their evolution in fluctuating environments is essential.

One additional way in which genetic diversity can be achieved in microbes is through hybridization. This is possible in microorganisms that can exist as diploids. Hybridization allows microbes to carry two versions of an allele, theoretically allowing for their survival in multiple conditions. In a study in which a *Saccharomyces cerevisiae* x *Saccharomyces uvarum* hybrids was created having two versions of the PHO8 gene, the high affinity phosphate transporter, it was observed that experimental evolution in low temperatures selected for the loss of heterozygosity and the maintenance of two copies of the cryotolerant parent (Smukowski Heil et al. 2019). These results suggest that whether the environment is static or fluctuating affects the outcomes of evolution and that hybridization may indeed be a rapid mechanism in which a microorganism generates diversity in its population to rapidly adapt to change.

## 1.5: Theoretical aspects of fluctuating selection

In the evolutionary framework, it is frequently stated that genetic variation is important for adapting to environmental change. This view is of course evolving as new studies identify new cellular levels at which phenotypic variation can be generated, however, the field of evolution is primarily focused on the study of allele frequency changes and as such many of the hypotheses are rooted in gene-based explanations. To understand the role of genetic variation in adaptation to change, fluctuating selection must first be defined. Fluctuating selection refers to a selective force that varies in

direction over a period of time. While this chapter has extensively described the effect of abiotic fluctuations on how cells respond, to comprehensively understand the principles that govern fluctuating selection, biotic factors will be considered as well. Frequency-dependent selection is a primary mode of fluctuating selection that changes in response to organismal influence. It occurs when the size of the population directly impacts the strength and direction of selection on the population; in other words it is when the rate of growth of the population is proportional to its size.

A classic example in population genetics of frequency-dependent selection is the predator-prey model based on hare and lynx populations that were observed to oscillate in coordination (Hewitt 1921). These observations were the basis for the Lotka-Volterra equations that predict oscillations in species abundance based on predator-prey interactions. Increases in predator frequency result in decreases in prey frequency which in turn result in starvation of the predator and decreasing its population size. This decrease in predator frequency then again results in a relaxation of selection on the prey and an increase in its population size. Microbial species recapitulate these interactive evolutionary dynamics in a number of ways.

Microbes are spatially connected and constantly in competition with each other and their frequencies can have important effects on neighboring cells. Competition can occur for environmental resources whereby an increase in frequency of one genotype can deplete resources for others. Another way in which microbes interact is through the release of metabolites that either negatively or positively affect the growth rate of

adjacent cells. An example of the exertion of negative effects on neighboring cells includes antibiotics that some fungal species are known to produce. This can also occur within species such as with toxic waste products such as ethanol in some bacterial species that limit their own growth after a certain density threshold. An example of positive effects include mutualistic interactions whereby two microbial species rely on the byproducts the other species produces as a nutritional source for growth and survival. These different types of interactions can result in fluctuating selection.

How these interactions affect biodiversity is of deep interest in microbiome studies aimed at understanding the determinants of community composition. One such example is portrayed in *E. coli* communities in which three distinct populations stably coexist over small spatial distances resulting in the maintenance of diversity within the community in what is known as rock-paper-scissors games. In this model, competition drives the oscillating dynamics of abundance in which each genotype transiently rises to the highest frequency until another genotype outcompetes it temporarily (Kerr et al. 2002).

A longstanding question in the field of evolutionary biology is how fluctuating selection impacts the trajectory and outcomes of evolution. Several hypotheses have been proposed to predict outcomes and explain observations. Theodosius Dobzhansky originally coined the ‘balance hypothesis’ which suggested that genetically polymorphic loci maintained their variation in a population by selecting for heterozygotes (Hedrick 2007). The term ‘balancing selection’ has evolved to generally be defined as any type of selection that maintains polymorphism in a population. The role of temporally fluctuating

environments in imposing balancing selection has become better understood through a variety of studies. Experimental and observational studies have shown that periodic fluctuations that occur as a result of seasonal changes or nutrient content can select for the coexistence of genotypes specialized to each of the experienced environments (New et al. 2014; Bergland et al. 2014) an ecological phenomenon known as the “Temporal Storage Effect” (Chesson 1994; Letten et al. 2018; Tan et al. 2017). In this scenario, it is frequently observed that specialists oscillate in frequency over the period of change as they encounter the preferred condition followed by the reversal of direction in selection when they encounter the less preferred condition. While it is frequently stated that genetic variation is important for the adaptation of a population to environmental change, the forces that maintain genetic diversity in a population are less well understood. These studies suggest that fluctuating conditions can play a role by imposing a form of balancing selection. By contrast, a neutralist view aims to explain the maintenance of variation in fluctuating environments by a combination of other processes such as the continual generation of mutation in a population and the introduction of variation through migration (Bertram and Masel 2019)Hedrik et al. 1976; (Bertram and Masel 2019).

## 1.6 Conclusion

Studying how fluctuating environments influence the dynamics and outcomes of evolution is essential for our basic understanding of adaptation as well as for a variety of applications. In an environment that is constantly in flux, understanding the imposed selection allows us to understand how modern-day organisms evolved. Additionally the

principles of how temporal changes determine outcomes in evolution can be used as tools. For example, *Pseudomonas aeruginosa*, a human pathogen, has been shown to respond differentially depending on the regime of antibiotic administration. Antibiotics given in a fluctuating regime were less likely to result in the evolution of resistance in *P. aeruginosa* than if the antibiotics were administered continuously (Melnyk et al. 2017). Other potential applications include developing regimes for evolving microbial strains for fermentation during beer brewing as well as for the production of synthetic materials.

Understanding the principles of evolution in fluctuating environments still requires extensive study as most of the conclusions thus far have been heavily based on either theoretical work or observational hypotheses from the natural world. Systematic experimental studies aimed at deciphering and quantifying the parameters that determine the outcomes of evolution are still lacking. This thesis aims to tackle some of the numerous questions about fluctuating environments by describing a series of highly controlled experimental studies and the significance of their results.

# Chapter 2: Fluctuating environments maintain genetic diversity through neutral fitness effects and balancing selection

This chapter is based on the research paper “Fluctuating environments maintain genetic diversity through neutral fitness effects and balancing selection” by Farah Abdul-Rahman, Daniel Tranchina and David Gresham, which has been published on bioRxiv and submitted to the journal Molecular Biology and Evolution. I generated all of the data for Figures 1, 2, 3, 4, and 5. I generated all supplementary figures presented here except for supplementary figure 1, which Daniel Tranchina performed the analysis for and generated. I collaborated with Daniel Tranchina on performing data analysis throughout the manuscript. I wrote the first draft of the manuscript text with edits from David Gresham and Daniel Tranchina.

## 2.1: Abstract

Genetic variation is the raw material upon which selection acts. The majority of environmental conditions change over time and therefore may result in variable selective effects. How temporally fluctuating environments impact the distribution of fitness effects and in turn population diversity is an unresolved question in evolutionary biology. Here, we employed continuous culturing using chemostats to establish environments that switch periodically between different nutrient limitations and compared the dynamics of selection to static conditions. We used the pooled *Saccharomyces cerevisiae* haploid gene deletion collection as a synthetic model for populations comprising thousands of unique genotypes. Using barcode sequencing (barseq), we find that static environments are uniquely characterized by a small number of high fitness genotypes that rapidly dominate the population leading to dramatic decreases in genetic diversity. By contrast, fluctuating environments are enriched in genotypes with neutral fitness effects and an absence of extreme fitness genotypes contributing to the maintenance of genetic diversity. We also identified a unique class of genotypes whose frequencies oscillate sinusoidally with a period matching the

environmental fluctuation. Oscillatory behavior corresponds to large differences in short term fitness that are not observed across long timescales pointing to the importance of balancing selection in maintaining genetic diversity in fluctuating environments. Our results are consistent with a high degree of environmental specificity in the distribution of fitness effects and the combined effects of reduced and balancing selection in maintaining genetic diversity in the presence of variable selection.

## 2.2: Introduction

The dynamics of adaptive evolution in genetically heterogeneous populations depend on the strength of selection and the distribution of fitness effects among genotypes (Bell 2008). How selective environmental conditions and genetic variation contribute to evolutionary dynamics is one of the central questions in evolutionary biology. In genetically heterogeneous populations the fitness of different genotypes varies and selection acts to increase the relative abundance of advantageous genotypes. In the simplest scenario, comprising a single fitness component (i.e. a single selective force) and single variant differences between genotypes, the distribution of fitness effects can reliably predict the dynamics of adaptive evolution. However, the impact of variable selective conditions, that result from fluctuations in the environment, on the distribution of fitness effects and the corresponding impact on genetic diversity is poorly understood.

In natural and engineered environments, organisms frequently encounter fluctuating selection in the form of physical or biotic factors (Bell 2010; Nguyen, Lara-Gutiérrez, and Stocker 2020). Fluctuations in environmental conditions may occur

with a regular period, such as diurnal fluctuations or aperiodically, such as with the random temperature variations that occur throughout the day. Periodic environmental fluctuations comprise an enormous range of timescales and patterns ranging from hours, as is the case with the gut microbiome (Schlomann and Parthasarathy 2019), to millenia, such as the timing between glacial periods. The prevalence of periodic fluctuations at different time scales in diverse environments suggests that our understanding of how evolution has shaped extant organisms and our ability to predict the future outcomes of adaptation requires understanding how organisms respond to environmental change.

The molecular basis of adaptation depends on the selective regime of the fluctuating environment. Regularly repeating and predictable fluctuations have been shown to select for anticipatory strategies such as a memory-dependent response (Razinkov et al. 2013; Lambert and Kussell 2014) or programmed oscillatory behavior (Johnson and Golden 2006). Conversely, fluctuations that occur at random intervals may favor strategies that don't rely on forecasting future environmental conditions, such as sense-and-response (Uschner and Klipp 2014) or bet-hedging strategies (Childs, Metcalf, and Rees 2010; Olofsson, Ripa, and Jonzén 2009). The frequency of environmental fluctuation with respect to generation time is also a key determinant of adaptive strategies; if the fluctuations are extremely rapid with respect to generation time, an organism may sense a time averaged condition, whereas extremely slow oscillations with respect to generation time may result in organisms experiencing effectively static conditions (Lin and Kussell 2016; Cvijović et al. 2015). Moderate

fluctuations that fall between the two extremes and with periods less than the generation time likely apply selective pressure on regulatory pathways requiring an organism to respond to environmental change on the individual level.

A variety of theoretical expectations underlie the selective dynamics of genotypes in fluctuating environments. Balancing selection, generally defined as any type of selection that maintains polymorphism in a population, can explain the maintenance of genetic diversity in temporally varying environments. For example, fluctuations with periods over multiple generations can select for the coexistence of genotypes specialized to each of the two environments (New et al. 2014; Bergland et al. 2014) an ecological phenomenon known as the “Temporal Storage Effect” (Chesson 1994; Letten et al. 2018; Tan et al. 2017). In the extreme case, antagonistic pleiotropy, in which an allele that is beneficial in one condition is deleterious in another, can manifest between the different phases of a periodically fluctuating environment. By contrast, a neutralist view aims to explain the maintenance of variation in fluctuating environments by a combination of other processes such as the continual generation of mutation in a population and the introduction of variation through migration Hedrik et al. 1976; (Bertram and Masel 2019). Theoretical analyses of fluctuating environments have suggested that the efficiency of selection can be reduced in variable environments (Cvijović et al. 2015). It has also been proposed that varying environments themselves can trigger increased mutation rate and thereby increase population diversity (Nelson and Masel 2018), or act as catalysts for evolution through more efficient phenotypic exploration (Kashtan, Noor, and Alon 2007).

Empirical approaches to studying selection in fluctuating environments present several challenges. In natural environments, experimental variables are difficult to control. Experimental evolution in a lab setting provides a potentially powerful approach and as such a number of studies have investigated the effect of fluctuating environments on adaptive evolution using experimental evolution of microbes (Cooper and Lenski 2010). In general, experimental microbial evolution studies have focused on genetically homogeneous populations and the effect of *de novo* mutation. However, a small number of studies have made use of genetically heterogeneous populations to address effects of environmental fluctuations (Salignon et al. 2018). To the best of our knowledge, all studies of microbial evolution reported to date have used batch culture, which requires serial passaging and population bottlenecking, adding additional variables to the study design. The precise control of selection in batch culture is also challenging because nutrient concentration changes continuously with population expansion even when the culture medium is unchanged (Li et al. 2018).

To study the effect of environmental fluctuations on the dynamics of adaptive evolution, we used the chemostat to establish different selective regimes and study their effect on population genetic diversity and the distribution of fitness effects in *Saccharomyces cerevisiae*. We used synthetic populations consisting of the pooled nonessential haploid gene deletion library (~4,000 unique genotypes) and quantified population composition using barcode sequencing (barseq). We characterized fluctuating environments in the chemostat using both experimental studies and

mathematical modeling. We find that environments, in which nutrient concentration fluctuates, maintain greater genetic diversity than static environments. Increased genetic diversity in fluctuating environments results from the absence of a small number of highly fit and specialized genotypes that rapidly dominate populations evolving in static conditions and an enrichment in fluctuating environments of genotypes with neutral fitness effects. Many genotypes show non-linear and non-monotonic responses (log abundance versus time) to both static and fluctuating selection, but fluctuating environments uniquely select for a class of genotypes with oscillatory growth behavior. Oscillatory behavior contributes to large short term fitness effects that are not observed over the long term. Our study highlights the importance of environmental variability in evolutionary dynamics and provides a framework for modeling evolutionary scenarios that better reflect natural environments.

## 2.3: Results

The empirical study of adaptive evolution requires consideration of both the selective conditions and the heritable variation in a population. In this study, we combined continuous culturing using chemostats and the *Saccharomyces cerevisiae* haploid non essential gene deletion collection to study the effect of temporally fluctuating selection on standing genetic variation.

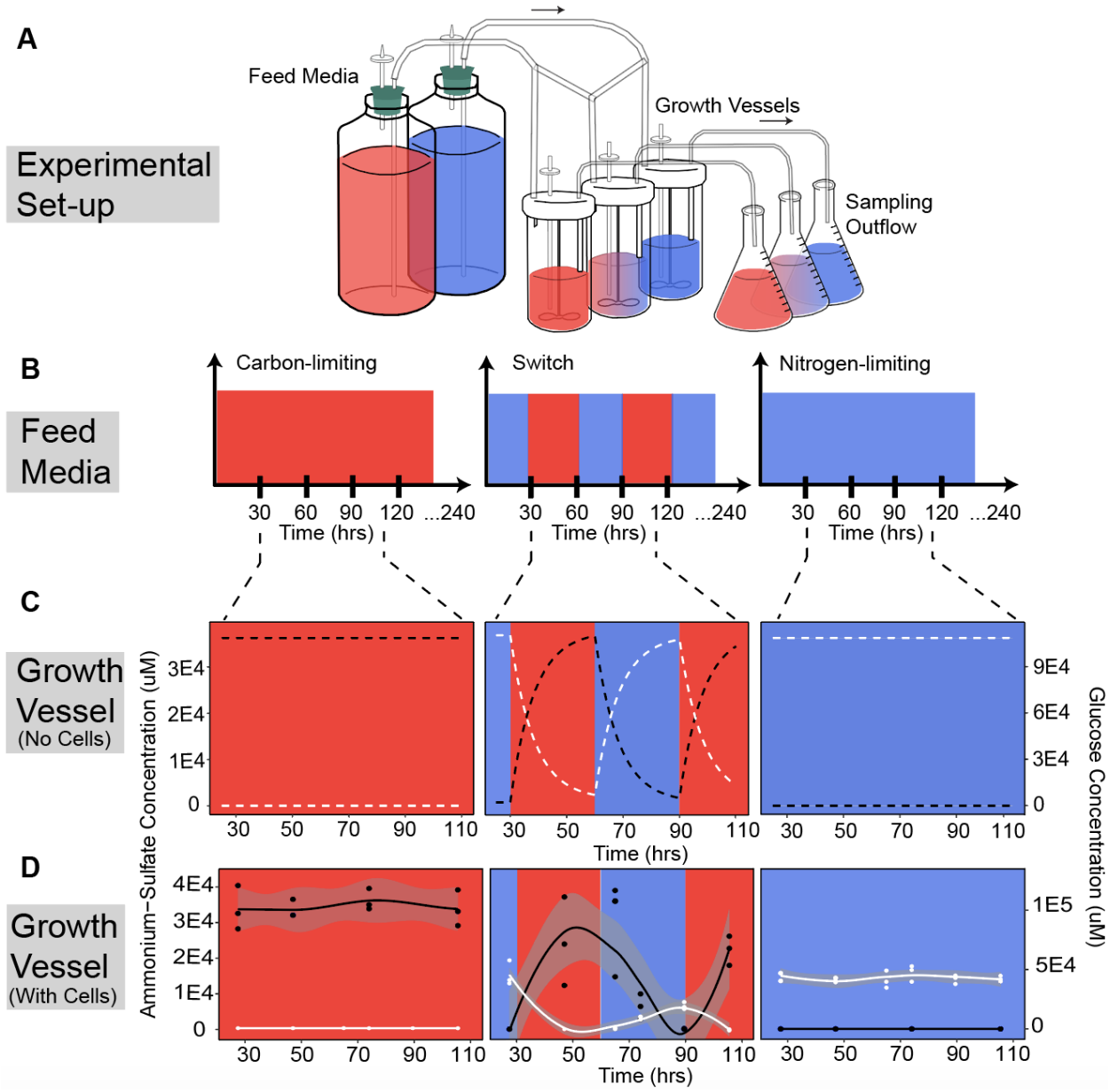


**Figure 2.3.1. Periodically fluctuating environments in the chemostat.** (**A**) We used chemostat cultures to study evolutionary dynamics in static and fluctuating conditions. (**B**) Populations were cultured in either carbon-limited media, nitrogen-limited media, or media that switched between the two nutrient limiting conditions every 30 hours (i.e. a period of 60 hours). All selections were maintained for a total of 240 hours. (**C**) An ordinary differential equation model was used to determine the expected concentrations of glucose (white), the sole carbon source, and ammonium sulfate (black), the sole nitrogen source, in the culture vessels in the absence of cellular consumption. (**D**) We experimentally measured glucose (white) and ammonium sulfate (black) concentrations in each of the culture vessels to determine the contribution of cellular consumption to environmental nutrient concentrations.

## 2.3.1 Modeling nutritional fluctuations in chemostats

Chemostats operate through continuous addition of fresh medium and removal of culture at the same rate (**Figure 2.3.1A**). We focused on two static conditions, carbon limitation (low carbon, high nitrogen) using glucose as the sole carbon source and nitrogen limitation (high carbon, low nitrogen) using ammonium sulfate as the sole nitrogen source. We constructed a periodically fluctuating condition in which the medium was switched between the two nutrient limiting conditions (**Figure 2.3.1B**). In this switch condition, the feed media alternates between the carbon limiting and nitrogen limiting media every 30 hours and the change in the feed media is instantaneous. We used the standard chemostat model (Novick and Szilard 1950; Monod 1950) to predict changes in nutrient concentration for single-nutrient limitation. We extended this model (**methods**) to incorporate temporal fluctuations in nutrient concentration and constrained cellular growth with a parameter for a second nutrient (Boer et al. 2010) to account for both changes in carbon and nitrogen concentrations.

We first modelled nutrient concentrations in the chemostat in the absence of cells to study the effect of dilution alone. Whereas a single limiting nutrient results in a constant nutrient concentration, switching the media results in oscillations in nutrient concentration in the growth vessel that follow first-order (exponential) kinetics despite instantaneous switches in the feed media (**Figure 2.3.1C**). We then inoculated chemostats with a single wildtype genotype and measured ammonium-sulfate and glucose concentrations in each of the culture vessels during steady-state cellular growth to determine the effect of cellular consumption on nutrient concentration in the

chemostat (**Figure 2.3.1D**). As expected, in all cases cellular consumption results in reduced nutrient concentrations in the chemostat. In the switch condition we find that the ammonium sulfate concentration oscillates between maximal and minimal levels that are equivalent to those observed in the two static conditions. By contrast, the maximal glucose concentration in the switch condition is reduced compared to glucose levels observed in static nitrogen limitation once the oscillations commence. This suggests that cells that have been previously exposed to growth-limiting levels of glucose exhibit increased glucose consumption in glucose rich conditions compared with cells that have not previously experienced growth-limiting glucose concentrations. This memory-like response may be due to the sustained expression of high affinity glucose transporters, induced by exposure to growth limiting glucose concentrations in the previous phase, persisting into the ammonium-sulfate limited phase (Rintala et al. 2008; Buziol et al. 2008).

We also considered an additional type of fluctuating environmental condition that differs in frequency and form. This fluctuation consisted of growth in steady-state ammonium-sulfate limiting conditions to which a bolus of 80uM nitrogen, in the form of either ammonium-sulfate or glutamine, was provided every three hours. This minor environmental perturbation, which we refer to as a “pulse” has previously been employed in studying transcriptional responses to environmental perturbation (Ronen and Botstein 2006; Airoldi et al. 2016). Prior mathematical modeling of chemostat pulses indicates that they result in a transient perturbation of the environment that rapidly returns to the steady-state condition (Airoldi et al. 2016).

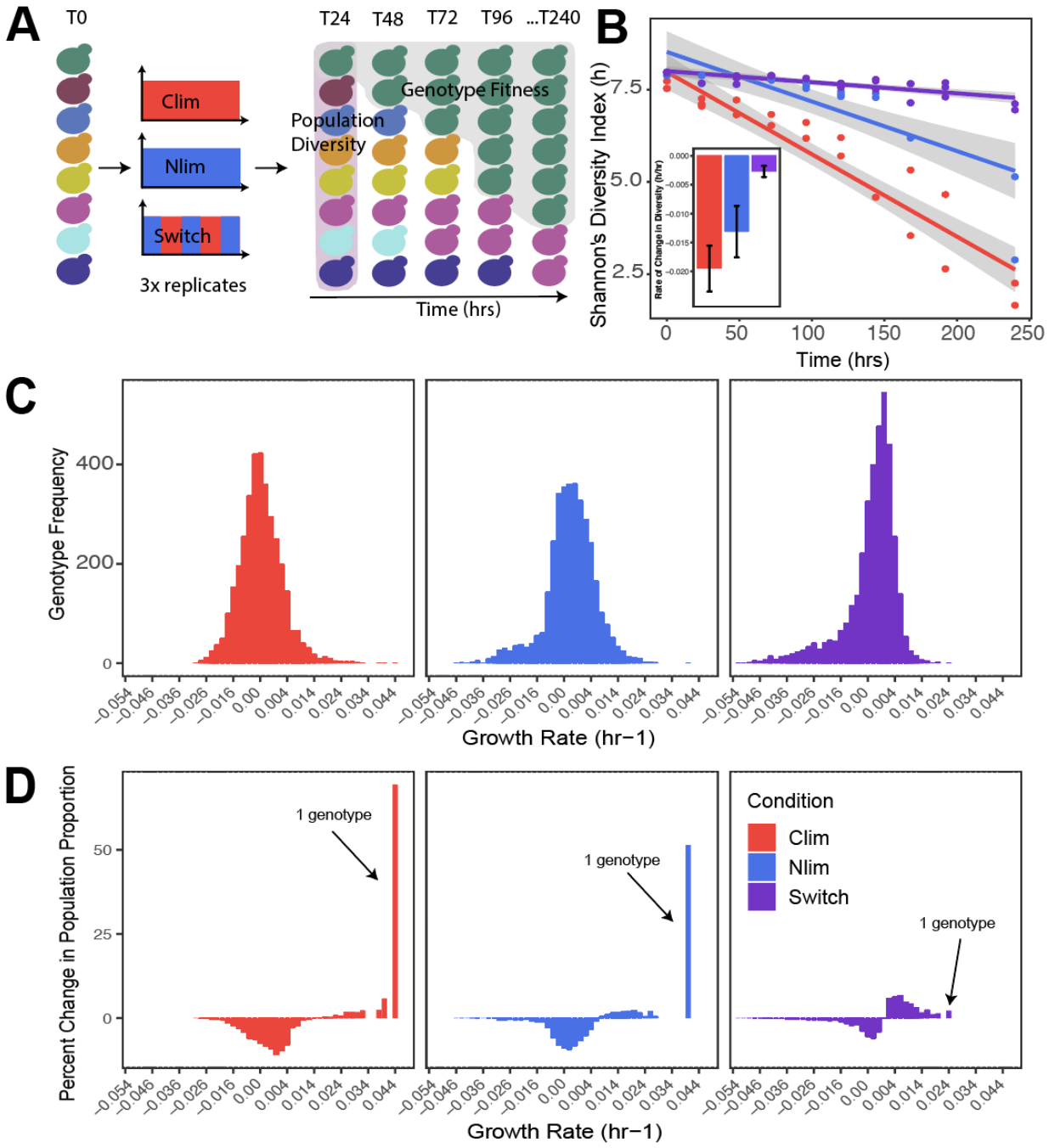


**Figure 2.3.2. Fluctuating selective conditions maintain greater genetic diversity than static selective conditions.** A single-gene deletion library containing ~4000 distinct gene knockout mutants was grown for 240 hours (approximately 40 generations) in static carbon-limitation, static nitrogen-limitation, and switching conditions in biological triplicate. Populations were sampled every 24 hours for a total of 10 timepoints. Barseq was used to estimate relative genotype abundance at each time point (**methods**). (**A**) Population diversity and genotype fitness were computed based

on changes in mutant abundance across time (**Appendix**). (**B**) The changes in Shannon's diversity index show a clear reduction in population diversity over time in static conditions, but not in a fluctuating environment. (**C**) The distribution of growth rates, relative to the arithmetic mean over all genotypes, for ~4000 genotypes in each condition estimated over the complete 240 hours of growth and (**D**) the change in the population proportion within each growth rate bin between t = 0 and t = 240 hours.

### 2.3.2 Selection on heterogeneous populations in a chemostat

We sought to quantify the dynamics of thousands of genotypes in static and fluctuating environments using chemostats. Classical chemostat theory holds that through the process of continuous growth and dilution, a chemostat attains a steady state in which the growing population size is constant and the concentration of the growth rate limiting nutrient is constant (Monod 1950; Novick and Szilard 1950; Kubitschek 1970). At steady-state, the population mean exponential growth-rate constant ($\lambda$) is equal to the chemostat dilution rate ($\beta$). However, competition for the limiting resource between the thousands of genotypes present in our experiments means that growth rates differ between genotypes. In our experiment, the growth rate of an individual genotype $i$, $\lambda_i$, is defined as the fitness of genotype $i$. Fitness differences across genotypes result in corresponding changes in population proportions over time. Intuitively, one might think that the changing proportions of genotypes would preclude constancy of the population growth rate. How can the constancy of population growth rate in the chemostat ($\lambda = \beta$) be reconciled with the presence of thousands of distinct genotypes with different fitness effects?

To address this question we modelled the growth of 4,000 genotypes in a nutrient-limited chemostat based on a straightforward extension of the two-genotype

model of competitive growth in a chemostat from (Dean 2005) (**Table 2.6.1**). As with Dean's two-genotype model, we find that the total population size and concentration of the limiting nutrient do in fact change as selection acts on the different genotypes. However, these changes are negligible after an initial transient period. We find that in the case of a static environmental selection in the chemostat, the genotype proportions change until a steady state is ultimately achieved in which only a single growth rate constant remains in the chemostat. In this new steady-state condition the population size is increased and the growth limiting nutrient concentration is decreased relative to the initial conditions (**Supplemental figure 2.6.1**). As the preceding period during which selection takes place is not a true steady-state, we refer to the selection during this time period as occurring in quasi steady-state conditions.

We also applied the Price equation in the continuous form (Day et al. 2020) to this scenario and found that the population growth rate cannot be exactly constant until the overall steady state condition above is achieved (**Supplemental figure 2.6.1**). Examination of the Price equation shows that evolution of the population growth rate is driven by the variance of the growth rate and the rates of change of genotype fitness (**Table 2.6.2**).

### 2.3.3 Fluctuating environments maintain greater genetic diversity

The distribution of fitness effects (DFE) quantitatively describes the proportion of variants in a population that are advantageous, neutral or deleterious, compared to the arithmetic mean fitness of the population. The shape of the DFE is influenced by several

factors including the type of species, population size, and genome size (Eyre-Walker and Keightley 2007). Whereas both theoretical (Connallon and Clark 2015) and experimental (Hietpas et al. 2013; Cooper and Lenski 2010; Blundell et al. 2017) studies have investigated the effect of a variety of environments on the DFE, the effect of temporal environmental variation on the DFE remains largely unknown. Moreover, the consequences of variable selection on the maintenance of genetic diversity is poorly understood.

To address the effect of variable selection on the DFE and genetic diversity we used an isogenic single-gene deletion library to compare selection in static and fluctuating environments. The presence of unique molecular barcodes enables identification of ~4000 genotypes using quantitative DNA barcode sequencing (barseq) (Delneri 2010). We used the haploid gene deletion collection and barseq to quantify population diversity and genotype fitness over approximately forty generations (240 hours) of sustained selection (**Figure 2.3.2A**). By replicating selection experiments and limiting their duration, our approach minimizes the potential confounding effect of *de novo* mutations. Assuming a rate of $2.7 \times 10^{-3}$ mutations/genome/replication (Drake et al. 1998) we would expect 0.108 mutations/ genome over 40 generations. Consistent with this expectation, after filtering sequencing libraries (**Figures 2.6.2A and 2.6.2B**, **Table 2.6.1)**, replicate populations showed high within-condition correlation indicating that *de novo* mutations did not play a significant role in selection dynamics. A small number of replicate experiments with low correlation were excluded from further

analysis (**Supplemental figure 2.6.2C**). Following quality control, our study comprised 278 barseq libraries.

To test the effect of environmental variability on population diversity we estimated the normalized abundance of each genotype at each time point in each condition (**Appendix**). We quantified the temporal change per unit time (in hours) rather than per generation to enable direct comparison between conditions as population growth rates in fluctuating chemostats are not necessarily determined by the dilution rate as they are in static conditions. We quantified population diversity using Shannon's diversity index, which takes into consideration the richness of genotypes and the evenness of their abundances. We find that the static carbon-limiting and nitrogen-limiting conditions display rapid declines in diversity in comparison to the switch condition (**Figure 2.3.2B**). To test the generality of this conclusion we applied the same analysis to the two pulse conditions. In the presence of pulsed perturbations, populations also maintained greater genetic diversity across time suggesting that this feature may be generalizable to different frequencies and forms of environmental fluctuation (**Supplemental figure 2.6.3**). We found that diversity estimates are not significantly affected by barcode library size (pearson r = 0.106, p-value = 0.218) (**Supplemental figure 2.6.4**) excluding confounding effects of library size on diversity metrics. In addition, population richness does not appear to undergo large changes over time in any selection regime suggesting that differences in diversity are largely driven by changes in evenness (**Supplemental figure 2.6.5**). All selections resulted in some degree of genotype extinction, defined by

the absence of a genotype in a one or more terminal time points. We did not identify a common set of extinct genotypes across all conditions (**Supplemental figure 2.6.6)**.

To test if differences in the rate of change in genetic diversity are associated with differences in fitness effects we computed the DFE for each condition. To quantify fitness over a given time interval (t1, t2) we use the temporal mean growth rate per cell minus the arithmetic mean over all genotypes. This is given by the difference between the log of normalized abundance at the two time points divided by the time difference (**Appendix**). By using the chemostat, the population exponential growth rate constant is set at 0.12 $hr^{-1}$, which is equal to the population mean growth rate over all genotypes to the extent that the total number of cells in the chemostat remains constant (**Appendix**. We calculated average genotype fitness using the first (t = 0 hours) and last (t = 240 hours) time point. The DFEs in each condition have similar distributions characterized by a unimodal distribution centered around neutral relative fitness (**Figure 2.3.2C**). The DFE in all three conditions comprises primarily neutral genotypes with tails reflecting deleterious and beneficial genotypes relative to the mean population fitness. This property also holds for pulse fluctuations (**supplemental figure 2.6.7**). Whereas measures of dispersion for each DFE are similar between conditions, contrary to previous predictions (Connallon and Clark 2015), static conditions are distinguished by the presence of individual genotypes with extreme fitness effects (**Table 2.6.2**). Thus, the distinguishing feature of the DFE, calculated over the entire period of selection, in static populations is the occurrence of extreme high fitness genotypes that are not

observed in fluctuating selections. This observation is consistent with theoretical analysis using the Price equation (**Appendix**)

The presence of a single particularly high fitness genotype results in a distinct population composition following 240 hours of selection. In both static selection conditions, a single highly fit genotype comprises over 50% of the population at this terminal timepoint (**Figure 2.3.2D**). By contrast, the maximal frequency of the highest fitness genotype is only 3% in the switch condition (**Table 2.6.2)**. In pulse fluctuations, the increased frequency of a small number of genotypes in the populations is apparent; however, this effect is reduced compared with static conditions (**Figure 2.6.7** and **Table 2.6.2**). These results point to a clear causal relationship between the presence of a single high fitness genotype and a rapid reduction in genetic diversity in static environments in which a single dominant selective pressure acts.

To test the generality of our observations we analyzed the dataset of *Salingon et al.* (Salignon et al. 2018) who studied the single-gene deletion library in two types of fluctuating environments using serial batch cultures and bottlenecking. In one of the fluctuating conditions (temporal variation in methionine concentration) we observed the same trend as our study. However, in the case of environments that fluctuate in salt concentration we find the opposite trend (**Figure 2.6.8**). In this case, it is possible that specific gene deletions (e.g. *CIN5Δ/Δ, YOR028WΔ/Δ, SRFI1Δ/Δ)* are uniquely able to respond to the fluctuation in salt concentration. Alternatively, the distinct nature of the environmental change in our study, which changes gradually in the case of the switch or

transiently in the case of the pulse, compared with the instantaneous change of *Salingon et al*'s experimental design may be an important factor. This is consistent with prior work suggesting that gradually changing environments result in greater clonal interference than instantaneously changing environments in which mutations of large beneficial effect are more likely to fix early (Morley and Turner 2017).

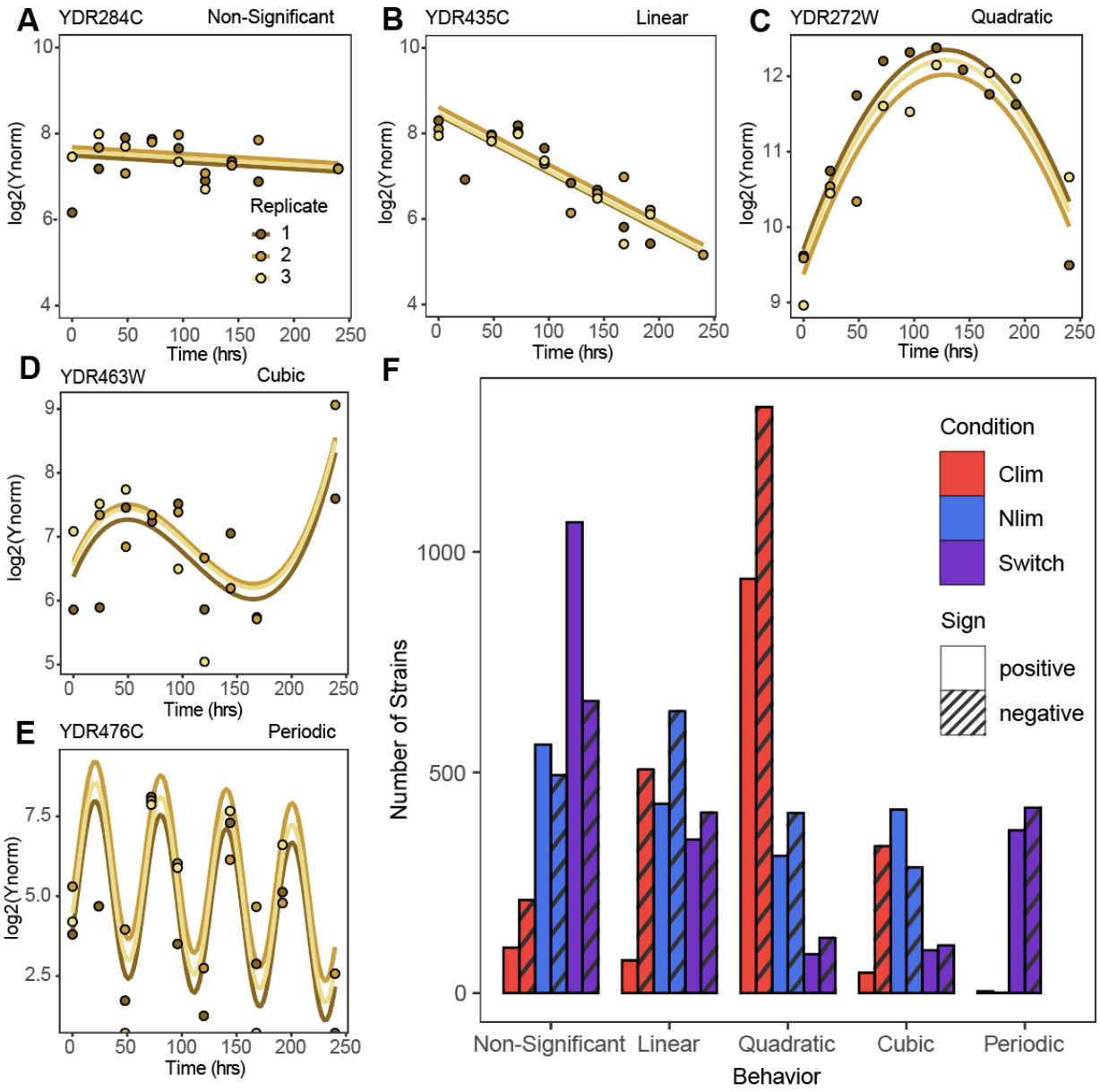


**Figure 2.3.3. Diverse temporal trajectories of genotypes in different selective conditions.** Temporal dynamics of genotypes across time fit to (**A**) non-significant, (**B**) linear, (**C**) quadratic, (**D**) cubic and (**E**) periodic models across time. (**F**) The distribution of temporal dynamics in each condition. Abundance changes are categorized as positive or negative based on the change in average growth rate between t = 0 and t = 240. Model fits for periodic models were defined as positive or negative on the basis of phase.

## 2.3.4 Genotypes exhibit distinct frequency dynamics

Whereas it has been widely demonstrated that cells exhibit rapid transcriptional (Spies et al. 2019; Airoldi et al. 2016; Ronen and Botstein 2006; Gasch et al. 2000) and physiological responses to changes in the environment (Bresson et al. 2020), the sensitivity of growth rate to environmental changes is less well studied. We sought to quantitatively describe the high resolution changes in genotype frequency across time for each genotype in each condition. The temporal dynamics of a genotype in a population is a result of both the externally imposed environmental selective pressure and the response to selection by all genotypes in the population. To characterize the dynamics of each genotype we performed hierarchical model fitting for each genotype using a model in which the log of the normalized barcode count from all ten time points is a polynomial function of time (**Appendix**). We explored the suitability of four distinct models of log normalized abundance versus time - linear, quadratic, cubic, and periodic. We quantified their applicability by starting with the most complex model and performing iterative model simplification using the log ratio of maximum likelihood test (**Appendix**).

We observed multiple distinct genotype dynamics. We find that between 10% - 30% of genotypes (**Figure 2.3.3F**) do not show a significant change in normalized abundance (**Figure 2.3.3A**) over the duration of the experiment. For these genotypes, the growth rates are not significantly different from the arithmetic mean over all genotypes. The prevalence of these genotypes is consistent with the greatest density of genotypes falling around a relative fitness of zero (**Figure 2.3.2C**). Although the mean relative growth rate is zero by our definition of relative growth rate, the concentration of

the distribution around zero relative growth rate is not guaranteed or predictable. Many genotypes show log-linear behavior across time (**Figure 2.3.3B**) indicating sustained positive or negative selection. Whereas static conditions in which selection is constant may be expected to result in such behavior, we find that almost a quarter of all genotypes also exhibit log linear behavior in the switch condition (**Figure 2.3.3F**). Such genotypes that are unaffected by alternations in the environment may be indicative of generalists. We identified non-monotonic genotype dynamics in all three conditions (**Figure 2.3.3F**). Quadratic behavior (**Figure 2.3.3C**) indicates an accelerating or decelerating growth rate per cell, whereas cubic (or sigmoidal) behavior (**Figure 2.3.3D**) reflects two reversals in the sign of fitness over the course of the experiment. A similar diversity of behaviors is found in the two pulse conditions (**Supplemental figure 2.6.9A**)

Our frequent sampling regime enables detection of genotype growth rate dynamics with high resolution. To that end, we tested whether genotypes show oscillatory behavior across the experiment (**Table 2.6.1**). We detect evidence for strong enrichment of periodically oscillating changes in genotype growth rate per cell (**Figure 2.3.3E**) that is unique to the switch condition (**Figure 2.3.3F** and **Supplemental figure 2.6.9A**). In these genotypes the growth rate per cell oscillates with a period that matches the period of environmental change imposed by switching the feed media to the chemostat. These genotypes include both positive and negative phases (i.e. with 180 degree difference). This behavior suggests a class of genotype that is acutely sensitive to variation in the environmental condition. To the best of our knowledge, there are few cases in which such oscillations in genotype frequencies have been observed.

One notable case is the oscillatory behavior of genotypes that has been observed over seasonal fluctuations in Drosophila populations (New et al. 2014; Bergland et al. 2014), Machado et al. 2018). In addition, high resolution sequencing of the 'Lenski lines' identified genotype oscillations in evolving *Escherichia coli* populations; however, this behavior eluded explanation (Good et al. 2017). Our finding suggests that such oscillations are potentially diagnostic of periodic variation in the environment. The 700 genotypes that comprise this class do not show significant enrichment for specific functions.

Non-monotonic behavior of genotypes may be the result of biological phenomena (e.g. environmental variation, genotype interactions, and density-dependent selection) or a consequence of data analysis methods. To test whether the highest frequency genotypes impact the apparent dynamics of other genotypes in the population, we computationally removed their barcodes from sequencing data, and repeated our complete analysis. We find that excluding the top performing genotype has a minimal effect on the identified non-monotonic growth behavior of the remaining genotypes (**Supplemental figure 2.6.9B**). As expected, the same manipulation has drastic effects on diversity metrics in static conditions, but only a minimal effect on the results observed for fluctuating conditions (**Supplemental figure 2.6.3**).

Fluctuating environments are enriched for genotypes that do not show a significant change in growth rate in comparison to other conditions (**Figure 2.3.3F** and **Supplemental figure 2.6.9A**). This observation along with the unique oscillating

genotypes point to two ways in which greater diversity is maintained in fluctuating conditions: 1) a larger fraction of genotypes have neutral fitness effects and 2) large fitness effects over short time spans undergo reversals in the direction of selection before they have a chance to dominate the population or go extinct.

### 2.3.5 Environmental fluctuations select for specific mutant classes

To identify the biological pathways and processes that contribute to increased fitness in each condition we performed gene set enrichment analysis (GSEA) (Subramanian et al. 2005) using the ranked fitness measurements for each condition. We find that constant carbon limitation selection results in the positive selection of gene deletion mutants with functions in carbon metabolism (**Supplemental figure 2.6.10**). The highest fitness genotype is deletion of *MTH1*, which has previously been reported as a target of selection in experimental evolution in glucose limited chemostats (Kvitek and Sherlock 2011). In static nitrogen limited chemostats, we find enrichment for genotypes with functions in nitrogen metabolism (**Supplemental figure 2.6.10**). The highest fitness genotype is deletion of *GAT1*, which we have previously identified as conferring a fitness advantage in ammonium-limited chemostats (Hong and Gresham 2014; Hong et al. 2018). Interestingly, in our previous studies we identified *GAT1* hypomorphs as beneficial, but *de novo* null mutations in *GAT1* were not identified.

We identified enrichment for distinct gene functions that are unique to the switch condition. Specifically, deletions in genes that encode components of the endoplasmic reticulum associated degradation (ERAD) pathway including *HRD1*, *HRD3*, *USA1*, and

*DER1* exhibit uniquely high fitness in the switch condition (**Supplemental figure 2.6.11**). The ERAD complex is responsible for degrading misfolded proteins in the endoplasmic reticulum (ER). Loss of ERAD function may be uniquely beneficial in fluctuating conditions as decreased rates of protein degradation may facilitate persistence of proteins across conditions thereby serving as a type of 'memory' response.

The periodic addition of excess nutrients in the pulse conditions results in the enrichment of unique classes of genotype function in addition to functions that are shared with the static conditions (**Supplemental figure 2.6.10**). This suggests that these transient environmental perturbations serve to both reduce the strength of selection and select for a unique class of genotypes.

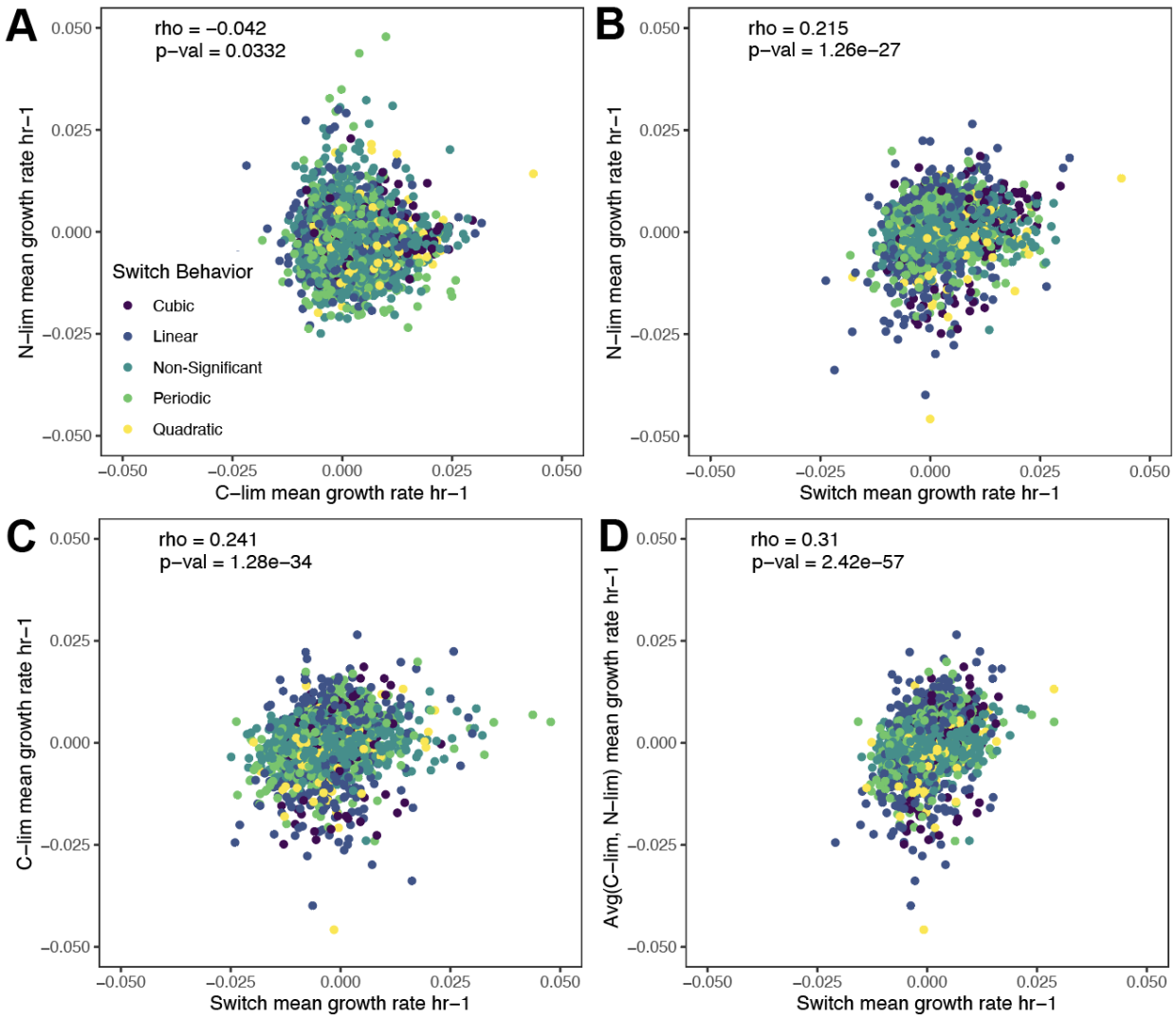


**Figure 2.3.4. A subset of genotypes have a predictive relationship between fluctuating and static selective conditions.** (**A**) The correlation in temporal mean growth rate per cell of genotypes between the two static conditions is low. There is intermediate correlation between the temporal mean growth rate per cell of the Switch condition and C-lim (**B**) and N-lim (**C**). The relationships between temporal mean growth rate per cell in the switch conditions and the average of the temporal mean growth rate per cell for the two static conditions has the highest correlation (**D**). Point colors indicate the model fit of the genotype as described in figure 3.

## 2.3.6 Fitness relationships between conditions

The fitness of a given genotype varies as a function of selection. We asked whether genotype behaviour under static selective conditions is predictive of fitness in a fluctuating environment. We find that the correlation between relative fitness in the two static conditions is low (**Figure 2.3.4A)**. The correlation between relative fitness in the carbon limited and switch condition (**Figure 2.3.4B)** and between the nitrogen limited and switch condition are somewhat higher (**Figure 2.3.4C**). We tested the simple model that fitness in a fluctuating environment is the mean of fitness in the two corresponding static conditions. We found that the correlation between the relative fitness in the switch condition and the mean of relative fitness in nitrogen limited and carbon limited conditions was only slightly increased compared with the correlation between each static condition and the switch condition fitness estimates (**Figure 2.3.4D**) .

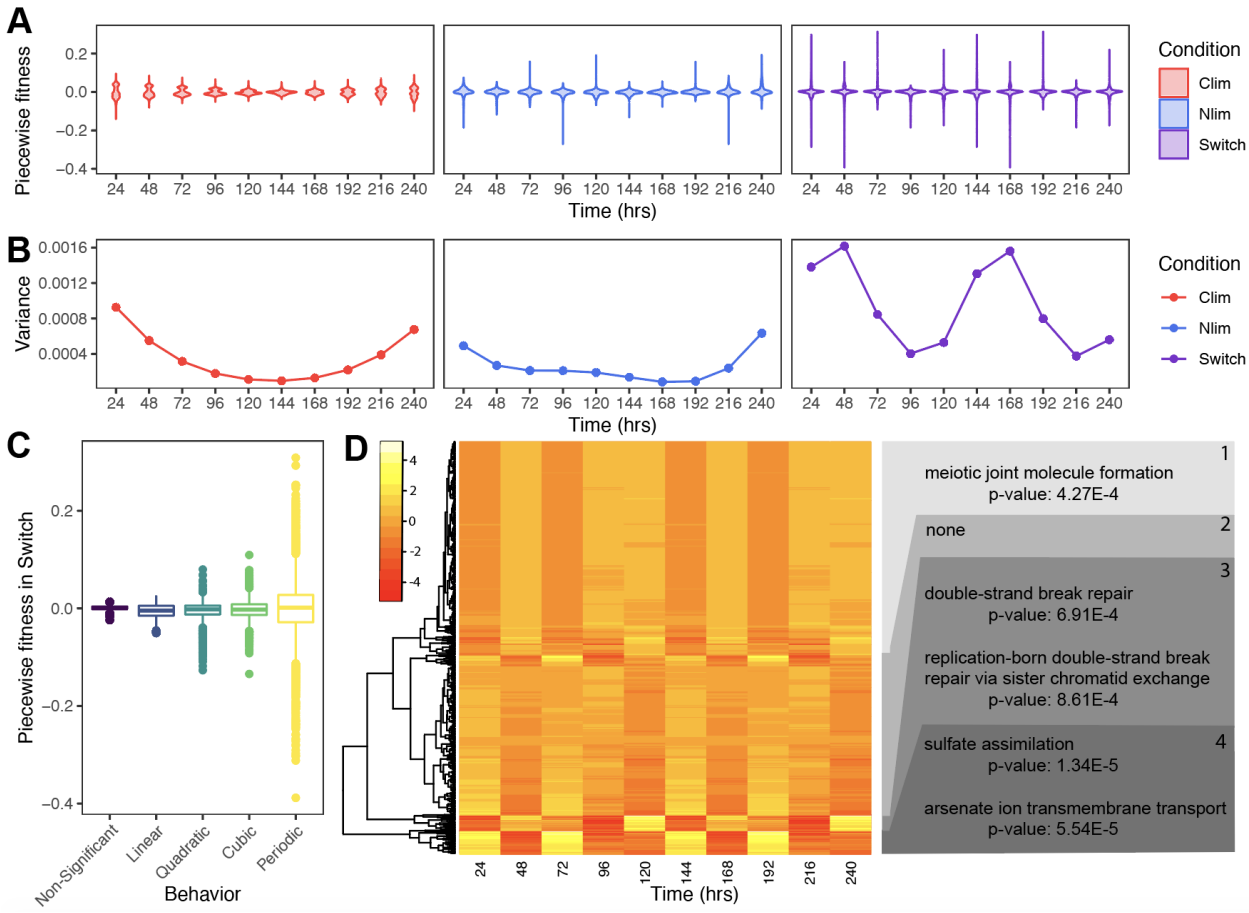


**Figure 2.3.5. The switch condition uniquely results in short term fitness changes that are not detected over larger timescales.** (**A**) Piecewise (temporal mean) relative fitness measurements were calculated by obtaining the difference between log normalized abundance at consecutive time points and dividing by the difference in time. Violin plots represent the distributions of piecewise fitness in each condition. (**B**) The variance of fitness measurements in each condition shows unique trends over time. (**C**) The distribution of piecewise fitness values according to best model fit. (**D**) Heatmap of scaled piecewise fitness for all periodically oscillating genotypes in the switch condition falling into four defined clusters. GO-terms that are enriched in each cluster are labeled on the right hand side.

## 2.3.6 Switching conditions harbor the highest short-term growth rates

To further understand how genotype behavior is affected in fluctuating conditions we compared short term fitness effects with long term fitness effects. Because we identified non-monotonic growth behavior, we calculated the piecewise fitness, defined

as the mean relative fitness values between consecutive time points, in the static and switch conditions. We find that whereas the temporal average relative fitness across the full time course shows minimal differences in DFE between conditions (**Figure 2.3.2C)**, the piecewise DFE is highly dynamic between timepoints and conditions (**Figure 2.3.5A**). Whereas static conditions select for genotypes with the highest average growth rate across the full time course, the switching environment results in the largest short-term fitness values. We computed the variance in fitness at each time point and found that static conditions have a unique U-shaped variance pattern in contrast with the switch condition which showed oscillating piecewise fitness variance (**Figure 2.3.5B**). The large differences in variance in fluctuating conditions are explained by the behavior of the periodically oscillating genotypes which have the highest piecewise fitness values across all growth behaviors (**Figure 2.3.5C**). Periodically oscillating genotypes are not a uniform group as we identified four clusters of genotype behaviors. Three of the four clusters have unique overrepresented GO-terms suggesting functional coherence among these genotypes (**Figure 2.3.5D**).

## 2.4: Discussion

In this study, we investigated the effect of fluctuating environments on genetic diversity and the distribution of fitness effects. We find that population diversity is greatly reduced in static environments compared with fluctuating environments. Our results support the idea that static environments impose stronger selection whereas fluctuating environments reduce the efficiency of selection (West and Mobilia 2020). By testing two distinct classes of environmental fluctuation we demonstrate that this result holds for two different types of environmental fluctuations.

We find that the maintenance of genetic diversity in fluctuating environments is a result of a combination of factors. First, genotypes with neutral fitness effects are enriched in fluctuating environments. Second, the presence of a unique class of genotypes that oscillate in frequency in fluctuating environments. Although this class includes genotypes with the highest and lowest short-term fitness effects the periodic reversal in the direction of selection ensures their maintenance at intermediate frequencies in the population, consistent with balancing selection. Third, the absence of genotypes with extreme long term fitness effects in fluctuating environments in contrast to static environments that are characterized by a single genotype with a large positive fitness effect that rapidly increases in frequency in the population. There has been considerable debate whether genetic diversity is primarily maintained through neutral fitness effects or through balancing selection (Hedrick et al. 1976). We have found that in the case in which new mutation does not occur, both balancing selection and neutral fitness effects contribute to the maintenance of genetic diversity in fluctuating environments.

Finally, we show that average fitness over long time spans can conceal the large variety of genotype behaviors in a population. Typically, fitness is estimated assuming monotonic behavior (Wiser and Lenski 2015) although a few studies have recently identified curvilinear dynamics (Schlecht et al. 2017). Our results suggest that the assumption of monotonic behavior is incorrect especially when considering population dynamics encompassing hundreds of unique genotypes, which is more representative

of dynamics in natural populations (Wiser and Lenski 2015; Landis et al. 2021). This is the case even in static selective conditions. More complex selective regimes that result from environmental fluctuations can result in more complex genotype dynamics as illustrated by the unique class of oscillating genotypes identified in our study.

## 2.5: Methods

### 2.5.1 Media

For all experiments, media consisted of defined minimal media supplemented with salts, metals, minerals, vitamins (Saldanha, Brauer, and Botstein 2004; Brauer et al. 2008; Airoldi et al. 2016). For glucose-limited media we added 0.08% glucose and 37mM ammonium-sulfate. For ammonium-sulfate-limited media we added 2% glucose and 400uM ammonium-sulfate. Static conditions used a single media source throughout the experiment. For the switch condition, we used a tube connecting the two feed media to a culture and alternated between the two media sources every 30 hours by manually clamping one inlet and opening the other. For pulse experiments we used the automated Sixfors chemostat system to deliver a bolus of either 40uM L-glutamine (PulseGln) or 40uM ammonium-sulfate (PulseAS) every three hours throughout the experiment.

### 2.5.2 Experimental measurements of model parameters

Measurements were taken at time points 2.5 prior to switch, then at 17, 35 , 44, 59.5, and 75.5 hours relative to the end of the first N-lim phase. This sampling scheme was chosen to capture the dynamics right after the first switch. Glucose was measured

using the r-Biopharm Glucose kit. Ammonia was measured using the QuantiFluo™ Ammonia/Ammonium Assay Kit. Cell density and cell size was measured using a Coulter Counter.

### 2.5.3 Mathematical modeling of chemostat growth in fluctuating environments

Population growth rate and the rate of change in the limiting nutrients glucose, and ammonium-sulfate were modeled using the following system of ordinary differential equations.

$$\frac{\partial S_{as}}{\partial t} = -S_{as} \cdot D + R_{as} \cdot D - \frac{X}{Y_{as}} \cdot u_{max} \cdot \frac{1}{\left(1 + \frac{K_{as}}{S_{as}} + \frac{K_g}{S_g}\right)} \quad (1)$$

$$\frac{\partial S_g}{\partial t} = -S_g \cdot D + R_g \cdot D - \frac{X}{Y_g} \cdot u_{max} \cdot \frac{1}{\left(1 + \frac{K_{as}}{S_{as}} + \frac{K_g}{S_g}\right)} \quad (2)$$

$$\frac{\partial X}{\partial t} = u_{max} \cdot X \cdot \frac{1}{\left(1 + \frac{K_{as}}{S_{as}} + \frac{K_g}{S_g}\right)} - D \cdot X \quad (3)$$

With the following parameters: D, the dilution rate of the culture (culture volumes/hr); X is the cell density (cells/mL), Y is the yield (cells/mL/mole of the limiting nutrient), $u_{max}$ is the maximal growth rate constant (hr-1), R is the concentration (uM) of the limiting nutrient in the medium, and S (uM) is the growth limiting nutrient concentration in the chemostat. Equation (1) describes the changes in ammonium-sulfate concentration over time. Equation (2) describes the change in glucose concentration over time. Equation (3) describes the change in cell density over time. To study the effect of our

experimental design for switching environments we performed numerical calculations with cell number (X) set to zero.

### 2.5.4 Culturing conditions

Library construction was performed as described in (Sun et al. 2020). An aliquot ($1.7x10^9$ cells/mL) of the pooled prototrophic gene deletion collection (VanderSluis et al. 2014) was thawed and 118uL were inoculated into triplicate chemostats with 200mL media for each condition. We estimate that this results in inoculation of the culture with 10^4 cells of each of the 10^3 genotypes. Cultures grew in r batch mode overnight at 30℃ to allow cells to reach high density (3E7 cells/mL). The first sample was collected and then the media feed pumps were turned on and tuned to a dilution rate of $0.12^{-hr}$ to switch cultures to continuous growth. Samples were collected every 24 hours by passive sampling from the chemostat outlet for a total of 240 hours. Cell pellets were stored in -80C in a cell storage solution (0.9M sorbitol, 0.1M EDTA, 0.1M Tris). DNA extractions were performed using the Hoffman Winston DNA prep (Hoffman and Winston 1987). PCR amplification of barcodes of each sample was performed by using a universal primer and an indexed primer (Robinson et al. 2013). The P5 illumina adapter was incorporated to all samples. Barseq libraries were sequenced on a 1x75 bp run on a Illumina NextSeq500.

### 2.5.5 Analysis of Barseq data

Barseq analysis was performed as previously described (Robinson et al. 2013). Briefly, barcode sequencing reads were matched to their corresponding genotypes

using Barnone . Reads with base pair mismatches greater than 0 were excluded from the analysis. Libraries with less than 100,000 total read counts were removed (**Supplemental figure 2.6.1A**). Uptags and downtags for each genotype were summed and genotypes with aggregate counts across all conditions with less than 1000 were also removed (**Supplemental figure 2.6.1B**). DEseq2 was used to normalize libraries (Love, Huber, and Anders 2014).

### 2.5.6 Mathematical modeling of genotype behavior

A detailed description of methods used for both data analysis and theoretical studies is provided in the supplemental methods. Throughout, we define the following terms:

- *Growth rate*: the change in population size between 2 time points, divided by time.
- *Instantaneous growth rate*: the derivative dn/dt.
- *Per capita (per cell) rate of change*: growth rate normalized by population size and accounted for by both cell divisions and cell death.
- *Per capita (per cell) growth rate*: same as per capita (per cell) rate of change where cell death is negligible.
- *Piecewise growth rate*: the growth rates between all consecutive timepoints based on the predicted values.

## 2.6 Supplementary figures

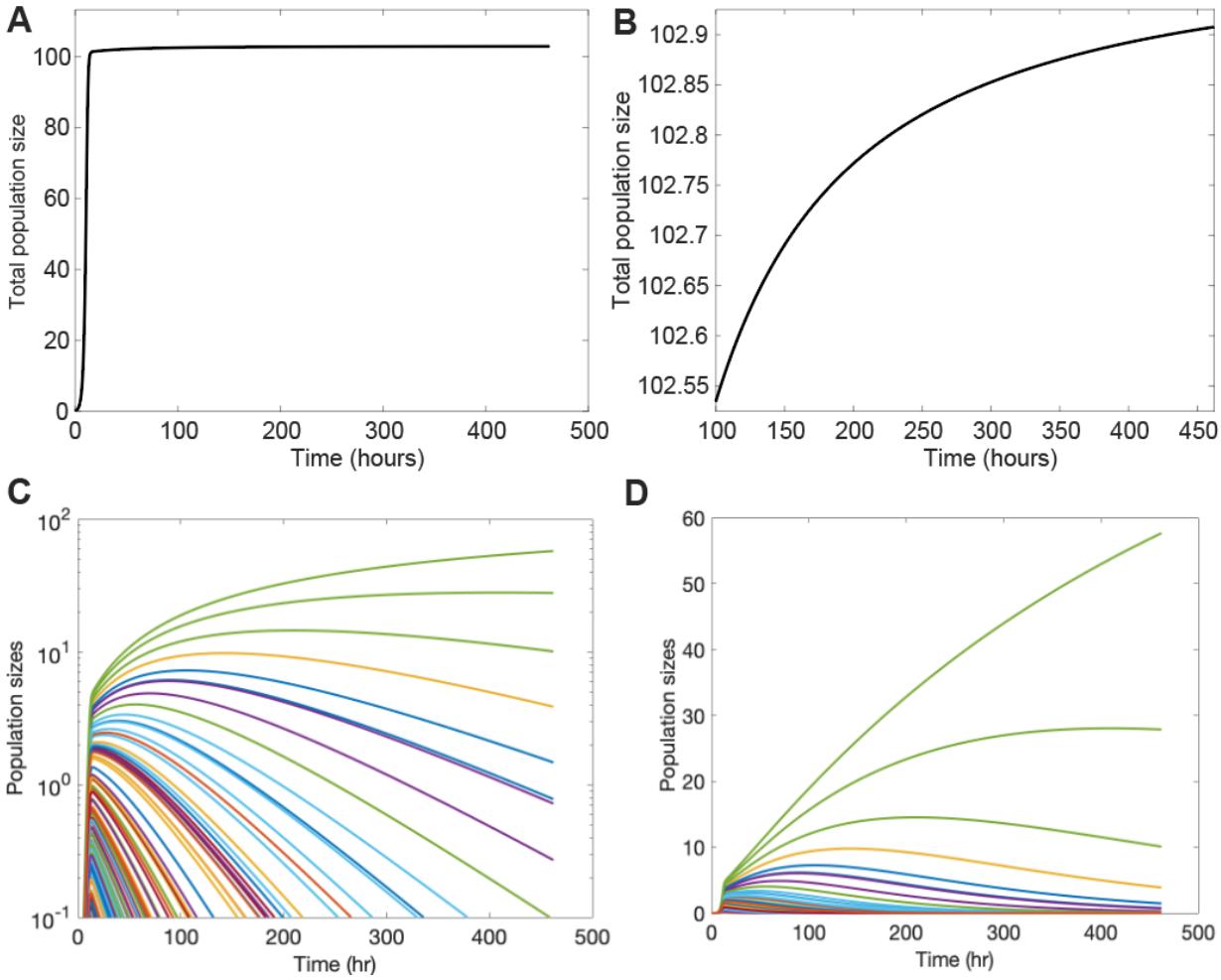


**Figure 2.6.1. Modelling the growth of four thousand genotypes in the chemostat. (A)** In the presence of thousands of genotypes the chemostat attains a quasi-steady state. **(B)** The total population size undergoes non-zero, but negligible, changes as selection acts on the population. **(C)** Individual genotype population sizes undergo large changes in frequency despite the relative invariance of total population size. (**D**) Dynamics of the top ten lineages.

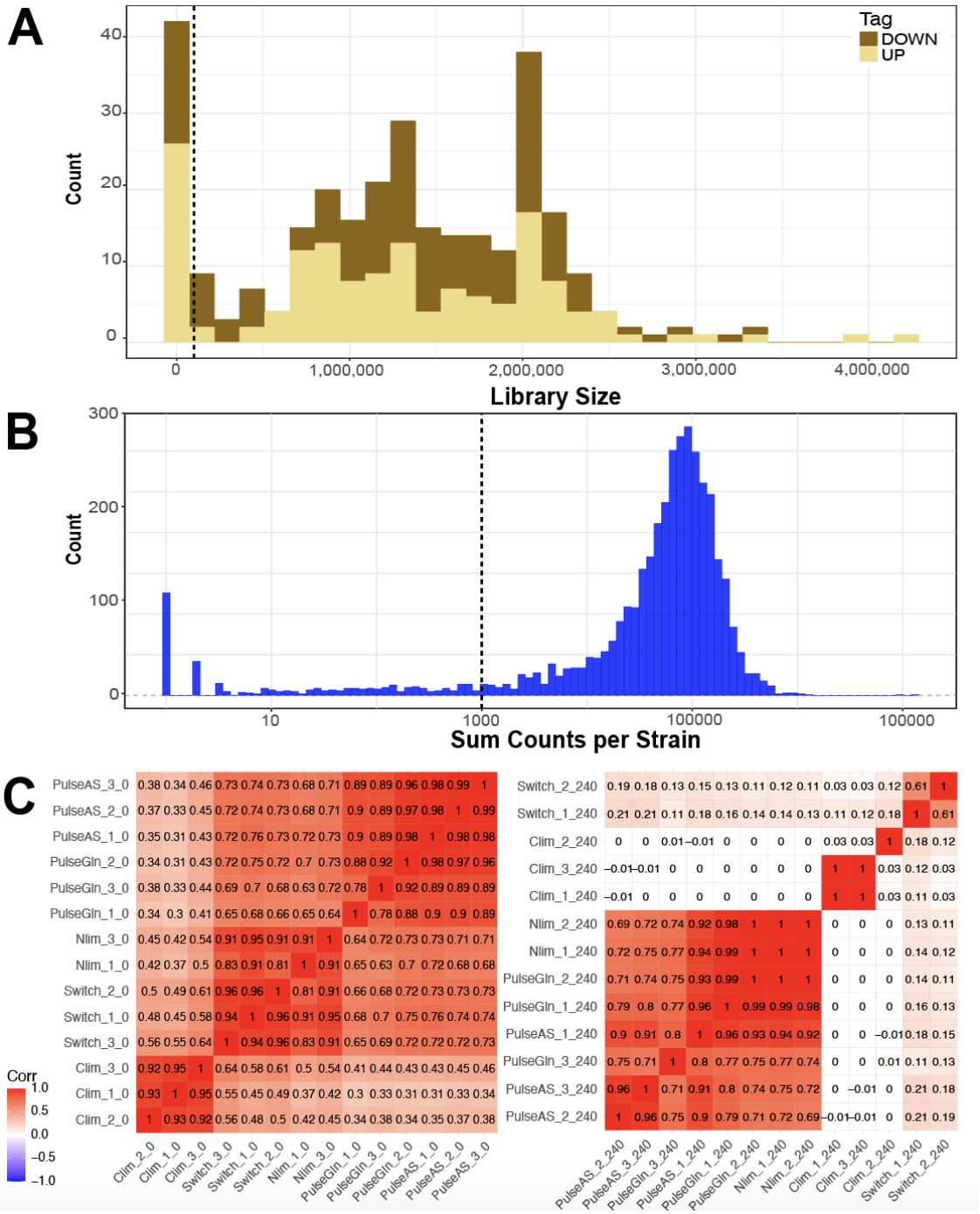


**Figure 2.6.2. Barseq library quality control.** (**A**) The complete distribution of library sizes before filtering is shown. The dashed line indicates libraries with less than 100,000 reads that were excluded from subsequent analysis. (**B**) The distribution of aggregate counts per genotype across all libraries. The dashed line indicates genotypes with less than 1,000 aggregate reads that were excluded from subsequent analyses. (**C**) Pairwise Pearson correlation coefficients between all samples at the first (t = 0) (**left panel)** and final time point (t = 240) (**right panel**). Replicates with correlation coefficients less than 0.6 were removed before proceeding with analysis.

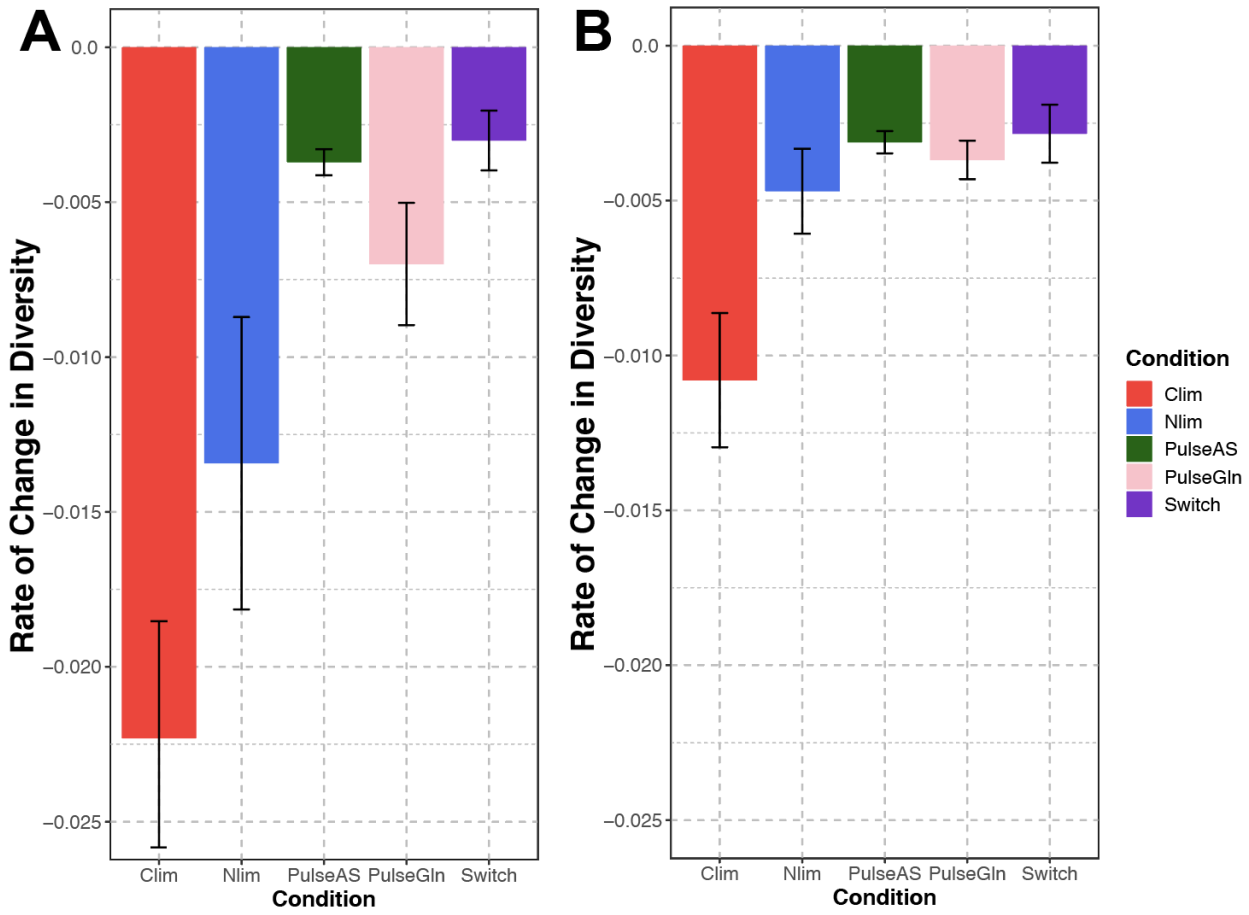


**Figure 2.6.3. Maintenance of genetic diversity in different fluctuating selections. (A)** Quantification of the dynamics of genetic diversity in other selective conditions. **(B)** Quantification of the dynamics of genetic diversity after excluding the highest fitness genotype in each condition.

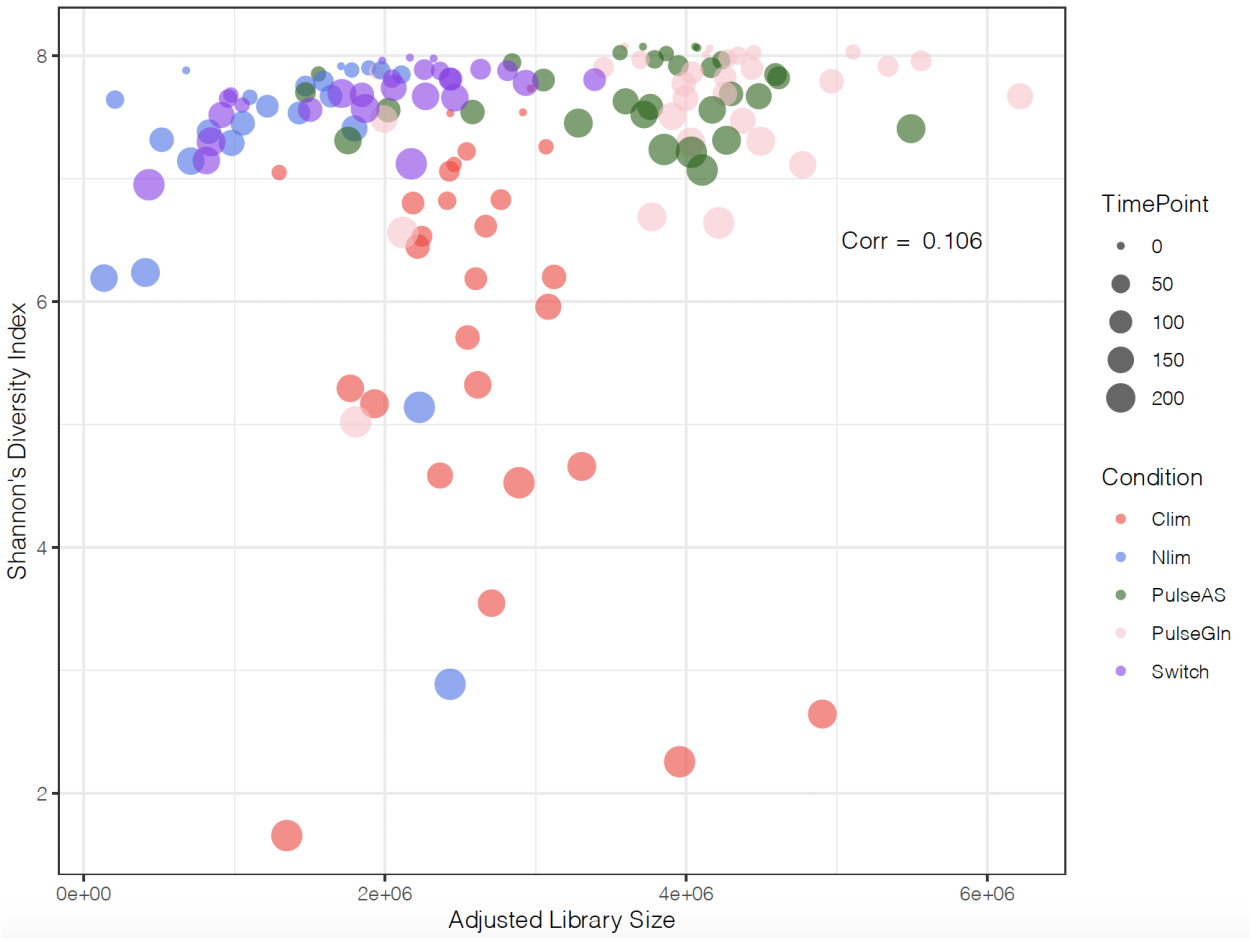


**Figure 2.6.4. Library size does not affect diversity estimates.** Plot showing the relationship between Shannon's diversity index and adjusted library size.

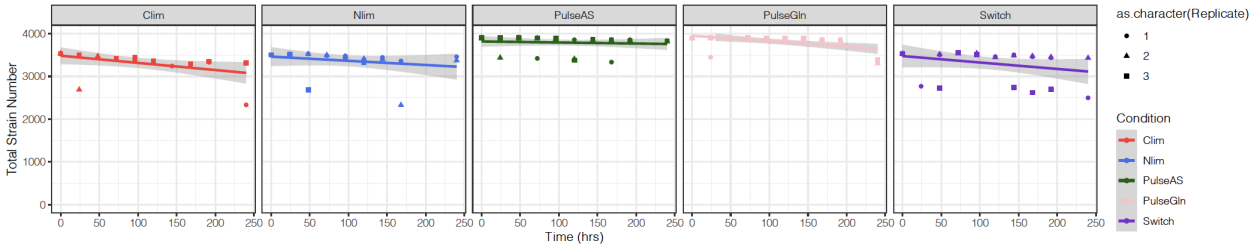


**Figure 2.6.5. The rate of change in total strain number across all conditions.**

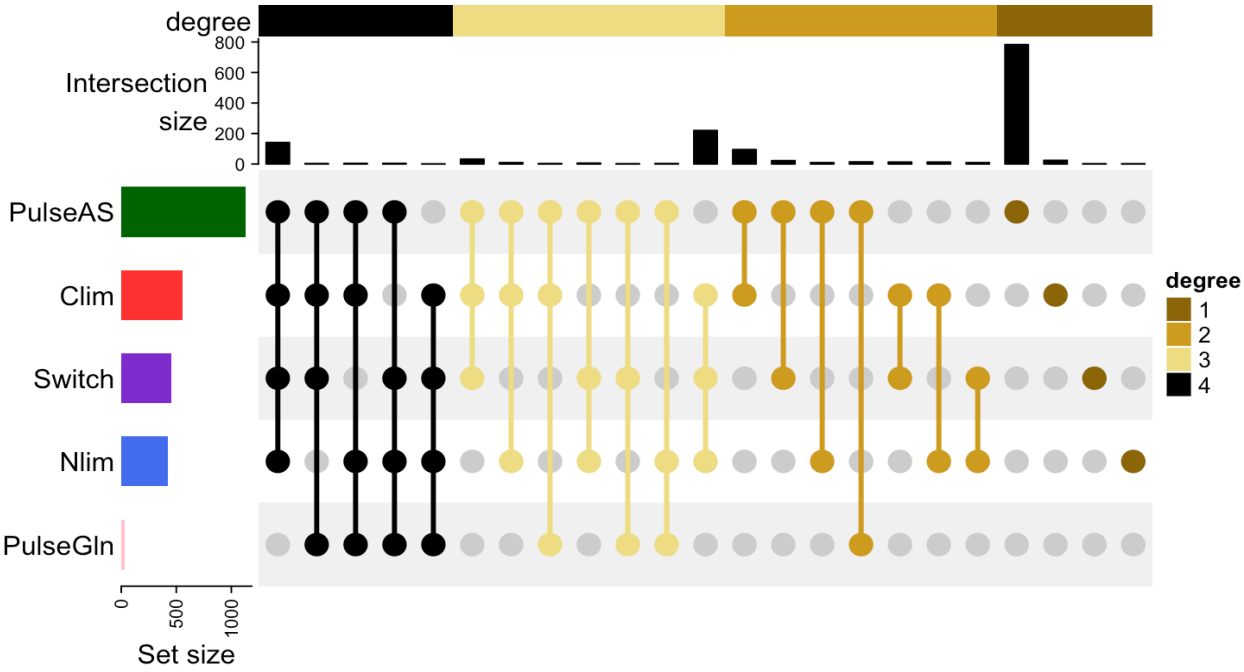


**Figure 2.6.6. Strain extinction profiles in each condition.** Extinct strains in the final time point (t = 240) are shared between subsets of conditions. Degree refers to the number of conditions that share a set of extinct genotypes.

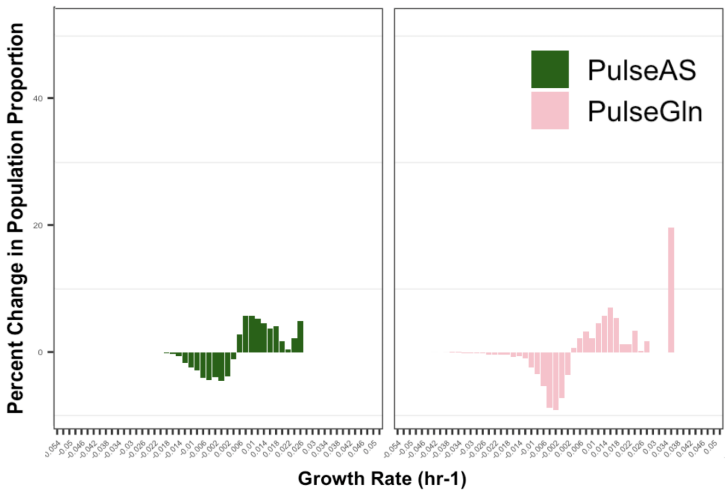


**Figure 2.6.7. Change in percent population proportion for pulse conditions.**

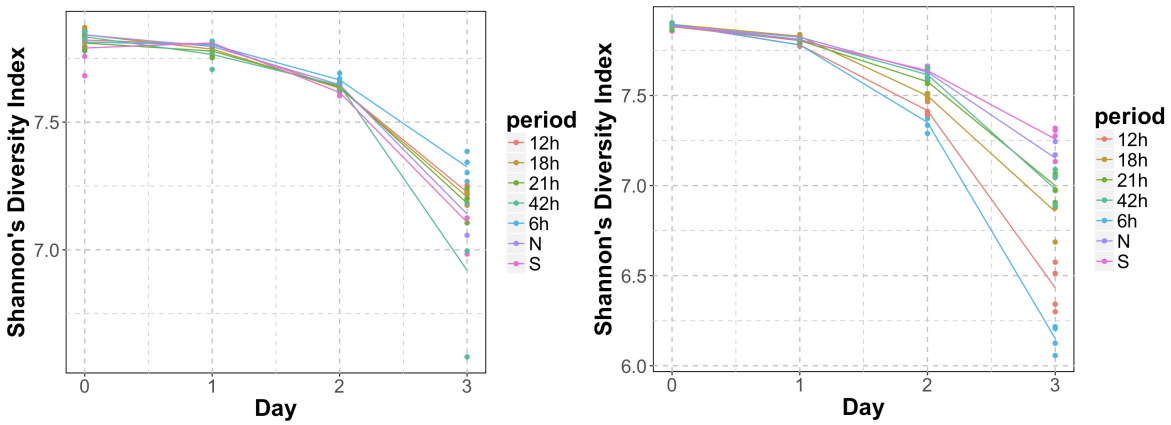


**Figure 2.6.8. Diversity measurements of experiments from the Salingnon et al. dataset.** Barseq was performed on the haploid gene deletion yeast library in conditions fluctuating between high (S) and low methionine (N) concentrations (left) and conditions fluctuating between salt (S) and no salt (N) concentrations (right).

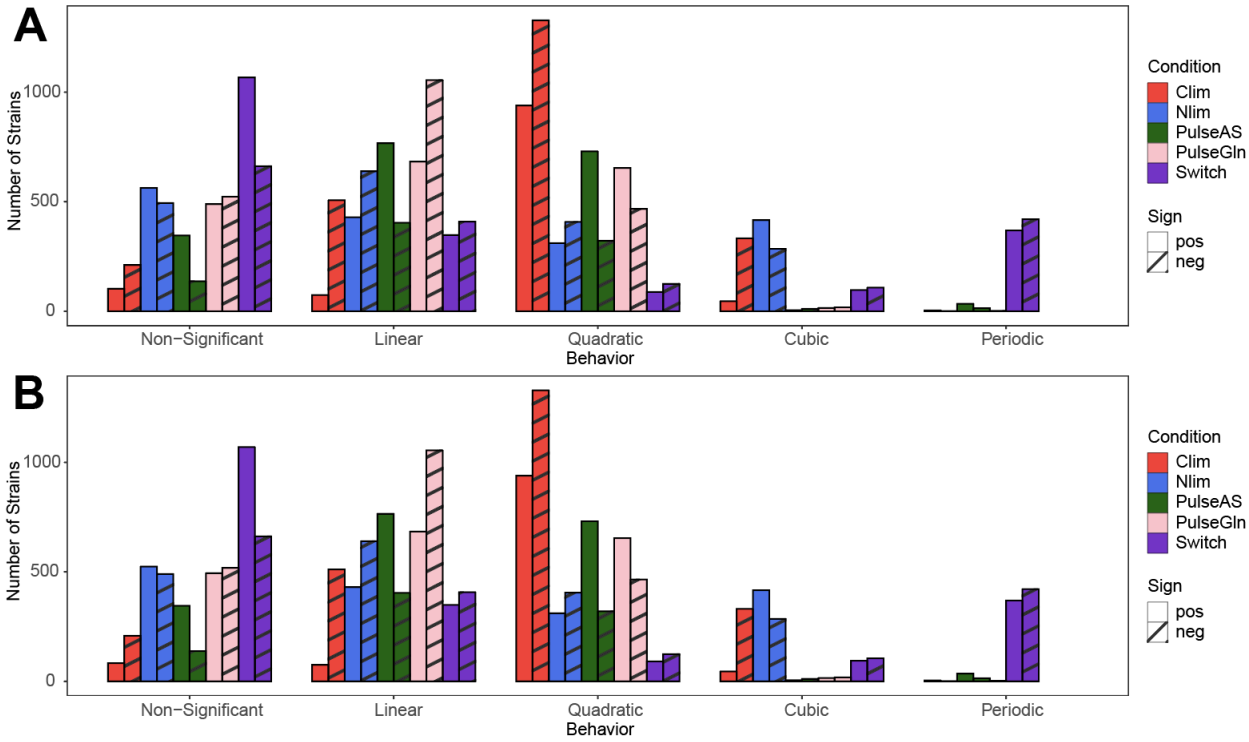


**Figure 2.6.9. Genotype dynamics in static and fluctuating environments. (A)** Extended summary of growth behavior including the two additional pulse conditions**. (B)** Reanalysis following removal of the highest fitness genotype does not alter the distribution of model fits**.**

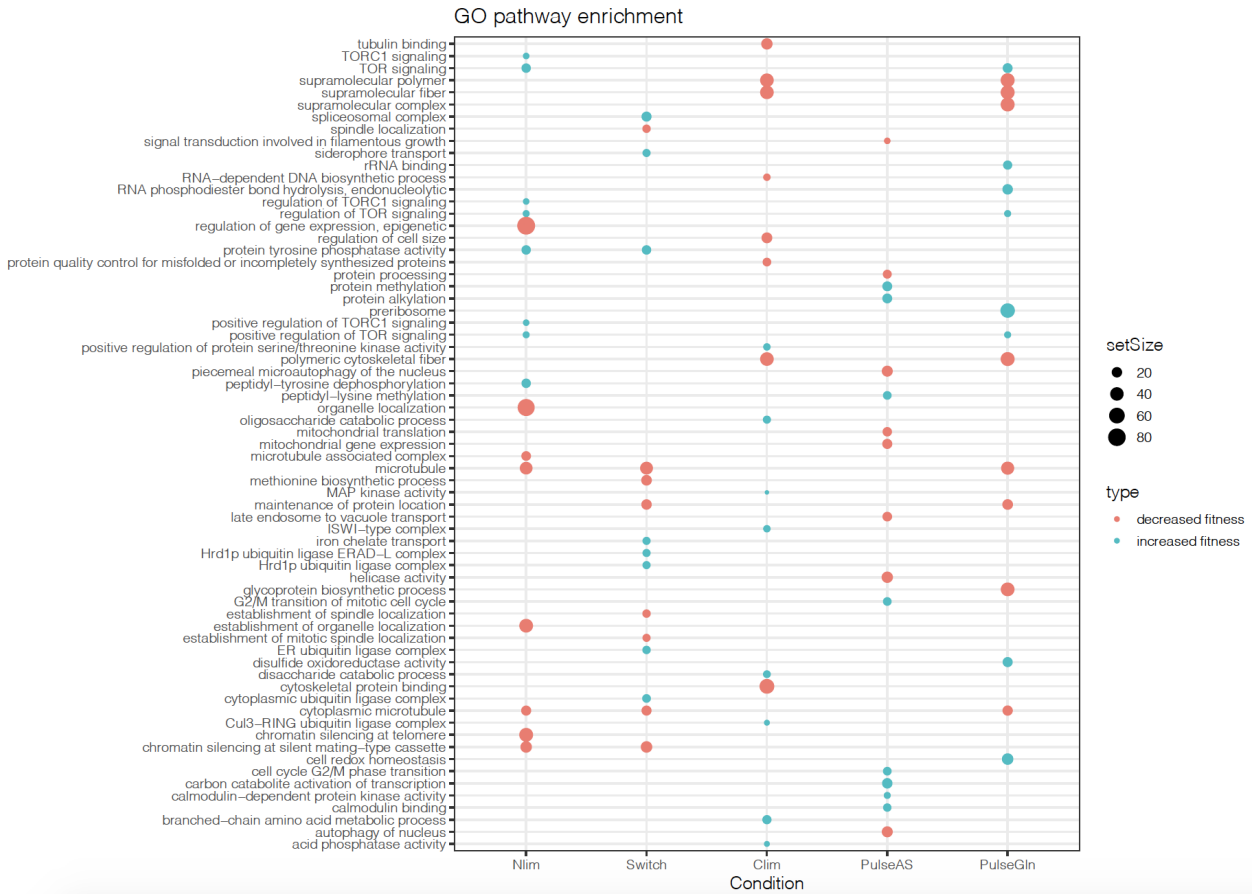


**Figure 2.6.10. Gene set enrichment analysis (GSEA) of fitness effects in each condition.** The top and bottom eight significantly (p-value < 0.05) represented GO terms for each condition are shown. Set size refers to the number of genes contained in a category.

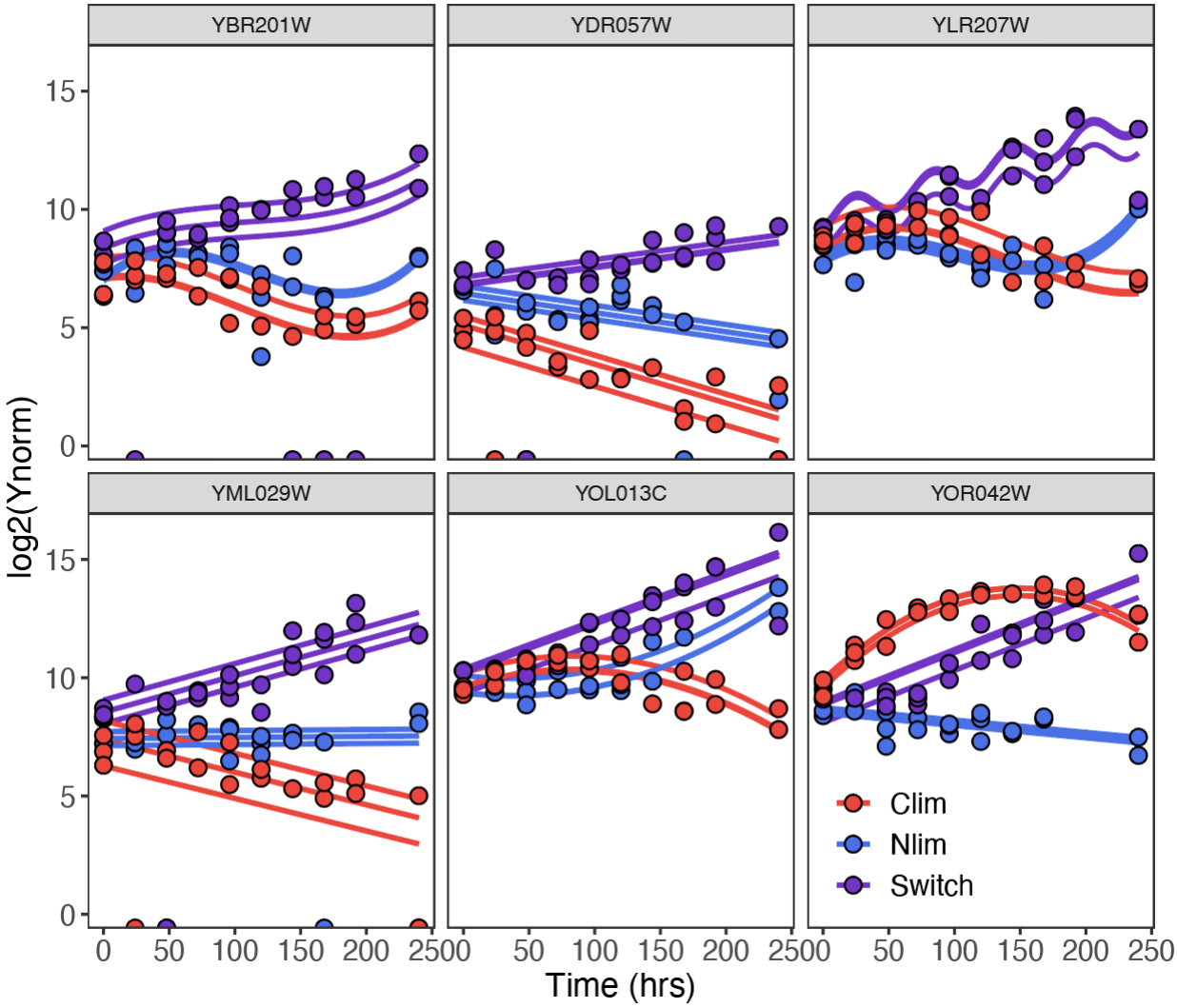


**Figure 2.6.11. Deletion of the ERAD genes uniquely results in increased fitness in fluctuating environments.** DER1, YOS9, HRD3, USA1, HRD1, and CUE5 gene deletions show consistent significant fitness increase in the switch condition but variable responses in carbon and nitrogen limiting conditions.

**Table 2.6.1. Pairwise correlation matrix of counts across all conditions**

**This table was not included in this thesis because it is a very large dataset. For access it can be found in the supplementary material for this manuscript posted on bioRxiv.**

**Table 2.6.2. DFE statistical measurements**

| Condition | Mode | Median | Mean | Max | Min | Range | Variance |
|---|---|---|---|---|---|---|---|
| Clim | 0.0077 | -0.0031 | -0.0027 | 0.0478 | -0.0271 | 0.0749 | 6.53E-05 |
| Nlim | 0.0124 | -0.0003 | -0.0007 | 0.0435 | -0.0439 | 0.0874 | 8.64E-05 |
| Switch | 0.0047 | 0.0005 | -0.0018 | 0.0265 | -0.0504 | 0.0769 | 9.53E-05 |
| PulseAS | 0.0020 | 0.0031 | 0.0027 | 0.0278 | -0.0176 | 0.0455 | 4.29E-05 |
| PulseGln | -0.0026 | -0.0004 | -0.0012 | 0.0361 | -0.0415 | 0.0776 | 6.94E-05 |

# Chapter 3: The dynamics of copy number variants in fluctuating environments.

This chapter is based on the research paper “The dynamics of copy number variants in fluctuating environments” by Farah Abdul-Rahman, Angela Hickey and David Gresham, which is in preparation. I generated all of the data for the Figures and Tables, wrote the manuscript text, and generated all of the supplementary figures presented here. Future and ongoing work generating data for this project is performed in collaboration with Angela Hickey.

## 3.1: Abstract

Microbes live in dynamic environments that pose ongoing challenges for reproduction and survival. Unlike static conditions in which a selective pressure is continuously experienced by the organism, fluctuating conditions cycle between different selection pressures resulting in complex evolutionary dynamics. Copy number variants (CNVs) are a class of mutation in which a genomic locus varies in repeat number. CNVs are widespread across all domains of life and have been implicated in diseases such as cancer. Despite the prevalence of fluctuating selection in natural environments, it is not well understood how variable selection influences the dynamics and diversity of CNVs in evolving populations. To study the generation and selection of CNVs, we evolved budding yeast populations over hundreds of generations in chemostats alternating between two conditions limited for nitrogen in the form of either glutamine or proline. Previously, we have found that genes encoding a glutamine transporter (*GAP1*) and proline transporter (*PUT4*) undergo duplication in chemostats limited for their respective nitrogen source. However, the evolutionary outcome of populations that undergo fluctuations between the two conditions is unknown. We developed a dual-fluorescence CNV reporter system for the two transporter genes

enabling inexpensive and rapid detection of duplications or deletions at the two loci simultaneously. We find that static conditions strongly select for transporter duplication with highly repeatable dynamics, whereas fluctuating environments result in greater variation between replicate populations and a lower frequency of CNVs at both loci. Our study highlights the impact of variable selection on evolutionary outcomes and suggests that CNVs are frequently associated with negative pleiotropy, which constrains their role in the absence of a single strong selective force.

## 3.2: Introduction

Changes in nutrient availability occur frequently in nature. Single-celled organisms are especially exposed to harsh fluctuations in the environment since they generally do not possess higher level structures and systems found in multicellular organisms that buffer transient perturbations. Fluctuations in nutrient availability can either be spatial such as through static chemical gradients, or temporal such as with seasonal changes. Changes in nutrient availability are widespread and can be the result of seasonal temperature change which influences the survival of plants, animals, and microbes. Additionally interspecies competition for available nutrients in non-motile organisms can result in boom and bust cycles, a phenomenon commonly observed in microbes.

Phenotypic variation is beneficial for populations that encounter environmental change by increasing the likelihood that an individual in that population expresses a phenotype that is beneficial in the new environment. Such variation can be generated either at the genetic level through DNA mutations or through variable gene expression

in genetically identical populations (Frank 2013). Genetic variation arises through the accumulation of mutations across several replication cycles, especially in instances in which selection is relaxed and a greater variety of mutations is maintained. This is generally considered to be a slower mechanism of generating phenotypic diversity since it relies on longer time lines which are dictated by the generation time of an organism (Matic 2019). Generation of heritable variation can be accelerated by increasing the mutation supply rate such as in hypermutator strains in *Escherichia coli* (Sniegowski, Gerrish, and Lenski 1997) in which mutation rate is under selection and methyl-mismatch repair genes such as *mutS* are inactivated resulting in an increased error rate per replication cycle (Rosche and Foster 1999; Kang et al. 2019). Non-heritable variation in expression, on the other hand, is the result of noise in gene expression resulting in stochastic behavior. Examples of this include stochastic switches in which isogenic individuals in a population turn on or turn off a gene with some probability, resulting in a phenotypically heterogeneous population.

Mutations are generated at different rates as a result of genomic context, location in the genome, and the molecular mechanisms responsible for generating them. Single nucleotide variants (SNVs) for example are the most frequently occurring class of mutation in the genome as they result from errors in DNA replication (Hong and Gresham 2014). Copy number variants (CNVs) are a class of mutation in which a genomic locus varies in repeat number either through duplication or deletion. There are a number of ways in which a single gene can be duplicated or deleted ranging from duplication of the gene itself to duplication of the entire chromosome or genome.

Chromosome or genome scale changes in copy number results from errors in disjunction during mitosis or meiosis. CNVs at individual loci can be generated by errors in recombination (Brewer et al. 2015). In addition, there are multiple replication-dependent mechanisms that generate CNVs including polymerase slippage, genomic rearrangement, and the reintegration of extracellular circles (Møller et al. 2015). Several DNA architectures can increase the likelihood of replication-dependent errors, such as long-terminal repeats, stretches of poly-A regions, and microhomology (Lauer et al. 2018). CNVs may lead to rapid and reversible genetic adaptations with large fitness effects that can be instrumental in adapting to recurring environmental change.

CNVs have been extensively studied in *Saccharomyces cerevisiae*. For example the ribosomal DNA locus which is composed of a repeating array of 100-200 copies of a ~9.1Kb repeat unit, can expand and shrink during replication cycles in response to environmental signals (Mansisidor et al. 2018) suggesting that CNVs facilitate rapid adaptation. The copper-inducible *CUP1* array of 2-20 copies of a ~2Kb repeat unit is another example in which environmental conditions appear to stimulate the formation of this adaptive class of variation (Salim et al. 2021). In static nutrient-limited chemostats, CNVs are a common adaptive response. Transporter genes that are known to undergo gene amplification in response to sustained nutrient stress include the sulfur transporter SUL1 (Payen et al. 2014), the ammonium-sulfate transceptor, MEP2, the proline transporter, PUT4, the urea transporter DUR1,2 and the amino acid transporter, GAP1 (Lauer et al. 2018; Hong and Gresham 2014). The amplification of these transporter

genes likely increases the number of expressed transporters providing a strong competitive advantage in low nutrient environments.

Gene amplifications can impose a fitness cost as they may result in increased gene expression burden, genomic instability or dosage imbalance of components of protein complexes. Although selective conditions may promote the maintenance of a particular CNV, once the selection is removed the CNV may be subject to strong negative selection resulting in its rapid loss from the populations. To study how variable selection impacts the dynamics with which *de novo* CNVs are generated and selected we studied CNV dynamics in conditions that periodically fluctuate between different selections.

## 3.3: Result

To study the dynamics of copy number variant (CNVs) generation and selection in fluctuating environments we combined continuous culturing using chemostats and a dual-fluorescence CNV reporter in *Saccharomyces cerevisiae*. By alternating between two different nitrogen-limiting conditions that each selects for gene duplication at a unique genetic locus encoding a specific nutrient transporter, we established a model for studying the effects of fluctuating selection on the structure, diversity and dynamics of CNV-mediated evolution.

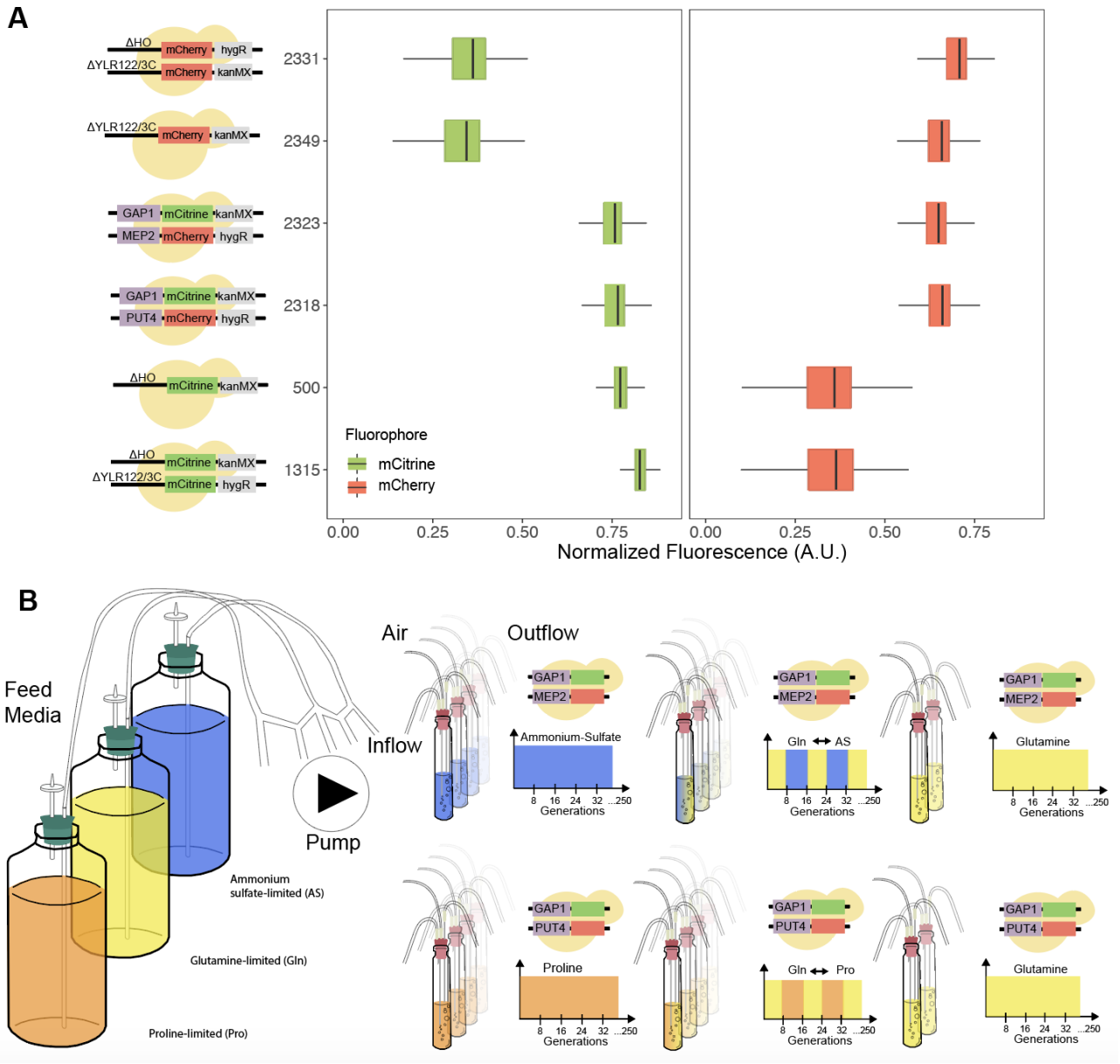


**Figure 3.3.1. The establishment of a model system for studying adaptive evolution of CNVs in fluctuating selection**. (A) Six strains were constructed with the constitutively expressed fluorescence reporter genes mCitrine and mCherry and the antibiotic markers kanMX and hygR. Two experimental strains were constructed in which mCitrine was inserted adjacent to *GAP1* and mCherry was inserted adjacent to either *MEP2* or *PUT4*. Four control strains were constructed with either one or two copies of mCitrine or mCherry inserted at neutral loci. The fluorescence profiles of each engineered strain is shown on the left for both fluorophores. Quantitative analysis of fluorescence confirmed that fluorescent gene copy number corresponds with fluorescence level. (B) Experimental evolution was performed using chemostats in continuous mode in replicates of six different conditions. Three conditions were inoculated with the *GAP1-MEP2* CNV reporter and three were inoculated with the *GAP1-PUT4* CNV reporter. Both strains were also inoculated in two separate vessels of

static glutamine-limitation. The *GAP1-MEP2* CNV reporter was inoculated in four vessels with static ammonium-sulfate-limitation, and four vessels that alternated between glutamine-limitation and ammonium-sulfate-limitation every eight generations. The *GAP1-PUT4* CNV reporter was inoculated in four vessels with static proline-limitation, and four vessels that alternated between glutamine-limitation and proline-limitation every eight generations. Experimental evolution was performed for 250 generations for all conditions.

### 3.3.1 Fluorescence is a proxy measurement for gene copy number at two distinct loci

We have previously developed a fluorescent CNV-reporter by inserting a constitutively expressed mCitrine gene adjacent to *GAP1*, and found single cell fluorescence to be a reliable reporter for changes in gene copy number during experimental evolution. Here, we extended this reporter system to allow simultaneous detection of CNVs at two genomic loci that are known to frequently undergo copy number variation. We constructed two dual-fluorescence reporter strains in which mCitrine was inserted adjacent to *GAP1*, and mCherry was inserted adjacent to either *MEP2*, or *PUT4*. We also constructed haploid control strains with either one copy or two copies of either mCitrine or mCherry inserted at neutral loci and found that copy number can be inferred on the basis of fluorescence for both CNV reporters (**Figure 3.3.1A**).

We performed experimental evolution in either static conditions limited for a single nitrogen source, or fluctuating conditions in which the media alternated periodically between two different types of nitrogen-limitation. The period of fluctuation was chosen based on the fact that many mechanisms that generate CNVs are DNA replication-based. Therefore, fluctuations that are greater than generation time are more likely to be meaningful intervals for the generation and selection of CNVs. For the

*GAP1-PUT4* CNV-reporter, the static conditions used were either glutamine-limitation or proline-limitation, and for the *GAP1-MEP2* CNV-reporter, the static conditions were either glutamine-limitation or ammonium-sulfate-limitation. The fluctuating environments for each genetic background alternated between the two respective static conditions every eight generations. All populations were evolved for a total of 250 generations (**Figure 3.3.1B**) with an effective population size of about $N_e \approx 3 \times 10^7$.

## 3.3.2 Periodic fluctuations in nitrogen quality do not alter cell size and density

Chemostats have been largely used for studying steady-state growth in static conditions. We have found that maintaining a constant limiting nitrogen concentration while alternating between nitrogen sources does not affect population density or size. Therefore, population growth rate, and therefore generation time, can be directly estimated from the culture dilution rate using standard chemostat theory. Using flow cytometry we also monitored forward scatter, a proxy for cell size, to determine whether fluctuations in nitrogen source result in changes in cell physiology. In instances when yeast cells are switched from growth on a less preferred nutrient source to a more preferred source, changes in cell size have been observed, however, in our experiments we did not find that cell size was significantly altered in fluctuating conditions (**Figure 3.7.2**).

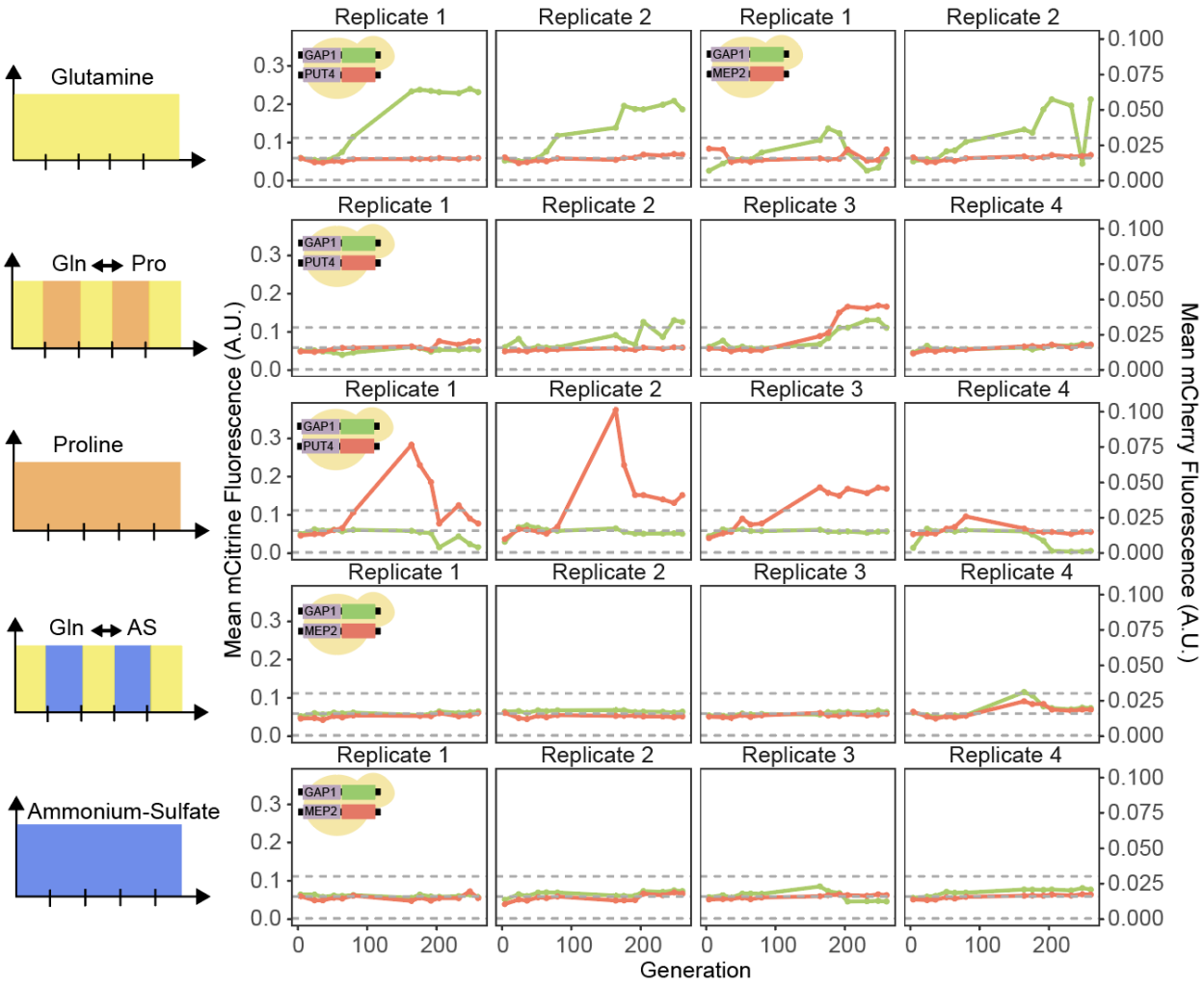


**Figure 3.3.2. A dual-fluorescence CNV-reporter system simultaneously tracks evolutionary dynamics of CNVs at two loci in static and fluctuating conditions**. Two dual-reporter strains were experimentally evolved in chemostats in 5 different conditions over 250 generations; three static conditions which were glutamine-limitation (Gln), proline-limitation (Pro), and ammonium-sulfate-limitation (AS), and two fluctuating conditions alternating between either Gln and Pro or Gln and AS. Two replicate populations were established with the *PUT4-GAP1* CNV reporter and two with the *MEP2-GAP1* CNV reporter. For all other conditions, the genetic background of the CNV reporter is indicated in the top left corner, and four replicate populations were established. The left-hand axis indicates the mean fluorescence level of mCitrine in each population and the right-hand axis indicates the mean fluorescence level of mCherry in each population. The horizontal dashed gray lines represent the mean fluorescence for the mCitrine and mCherry controls for 2-copy, 1-copy and 0-copies respectively from from top to bottom.

### 3.3.3 CNV evolutionary dynamics are condition- and genetic locus-dependent

Mean CNV reporter fluorescence for a population provides an overview of CNV dynamics in each condition for the three distinct genetic loci under study. As previously observed, *GAP1* duplications in glutamine-limitation predictably emerge and increase in frequency consistently across replicate populations as evidenced by the sharp increase in mean mCitrine population fluorescence starting around generation 70. This behavior occurs in four populations, two of which are founded by the *GAP1-PUT4* CNV reporter and the other two founded by the *GAP1-MEP2* CNV reporter suggesting that the addition of a second fluorescent CNV reporter at a second locus does not impact the dynamics of *GAP1* CNV evolution in an appreciable manner. In static proline-limitation, two populations undergo a decrease in mean mCitrine fluorescence to levels similar to control strains that have no mCitrine suggesting that *GAP1* deletions have increased in frequency. This is consistent with a trade-off between fitness in glutamine-limitation and fitness in proline-limitation potentially mediated by the presence of *GAP1*. In contrast, *GAP1* does not seem to have a similar fitness trade-off in ammonium-sulfate-limitation in which none of the populations undergo increases or decreases in mCitrine fluorescence. Assuming that the mutation rate of *GAP1* is constant between conditions, these results indicate that the selective pressure is the primary determinant of the observed dynamics.

In static ammonium-sulfate-limitation, the absence of increases in mean mCherry fluorescence suggests that *MEP2* does not undergo copy number increases at

appreciable frequencies. In contrast, in proline-limitation mean mCherry fluorescence increases indicating that *PUT4* increases in copy number. Mean fluorescence for both the *GAP1* reporter in glutamine-limitation and for *PUT4* in proline-limitation exceeds mean fluorescence levels of the 2-copy controls which suggests that increases in copy number greater than two-fold underly the observed dynamics. *PUT4* dynamics are unique in comparison to *GAP1* in that they show a sharp decline in copy number after generation 200, which could indicate that new mutations have arisen that confer greater fitness for cells in the population or that having such a high abundance of *PUT4* transporters in the population results in a depletion of environmental proline, resulting in a circumstance in which increased copy number of *PUT4* no longer confers a fitness advantage and is possibly deleterious. While the externally imposed environmental condition is static (i.e the concentration of proline is provided at an invariant rate), this may be an example of fluctuating selection that is mediated by the evolution in the population gene pool, which gradually shifts the selective pressure from the ability to scavenge nitrogen to pressure elsewhere, potentially towards metabolism or other processes downstream of nutrient import.

The relationship between fluctuating conditions and static conditions is complex. Environmental fluctuations do not simply result in an average of dynamics of the two corresponding static conditions. For example, in one population fluctuating between glutamine-limitation and proline-limitation, the *GAP1* mean fluorescence of the reporter oscillates in frequency, whereas in other populations it appears as though the fluctuation reduces the rate at which CNVs increase in frequency. In some populations undergoing

fluctuating selection there is no apparent increase in *GAP1* CNVs. Similarly, fluctuating between glutamine- and ammonium-sulfate limitation results in either transient increases in *GAP1* amplifications or a complete absence of amplifications.

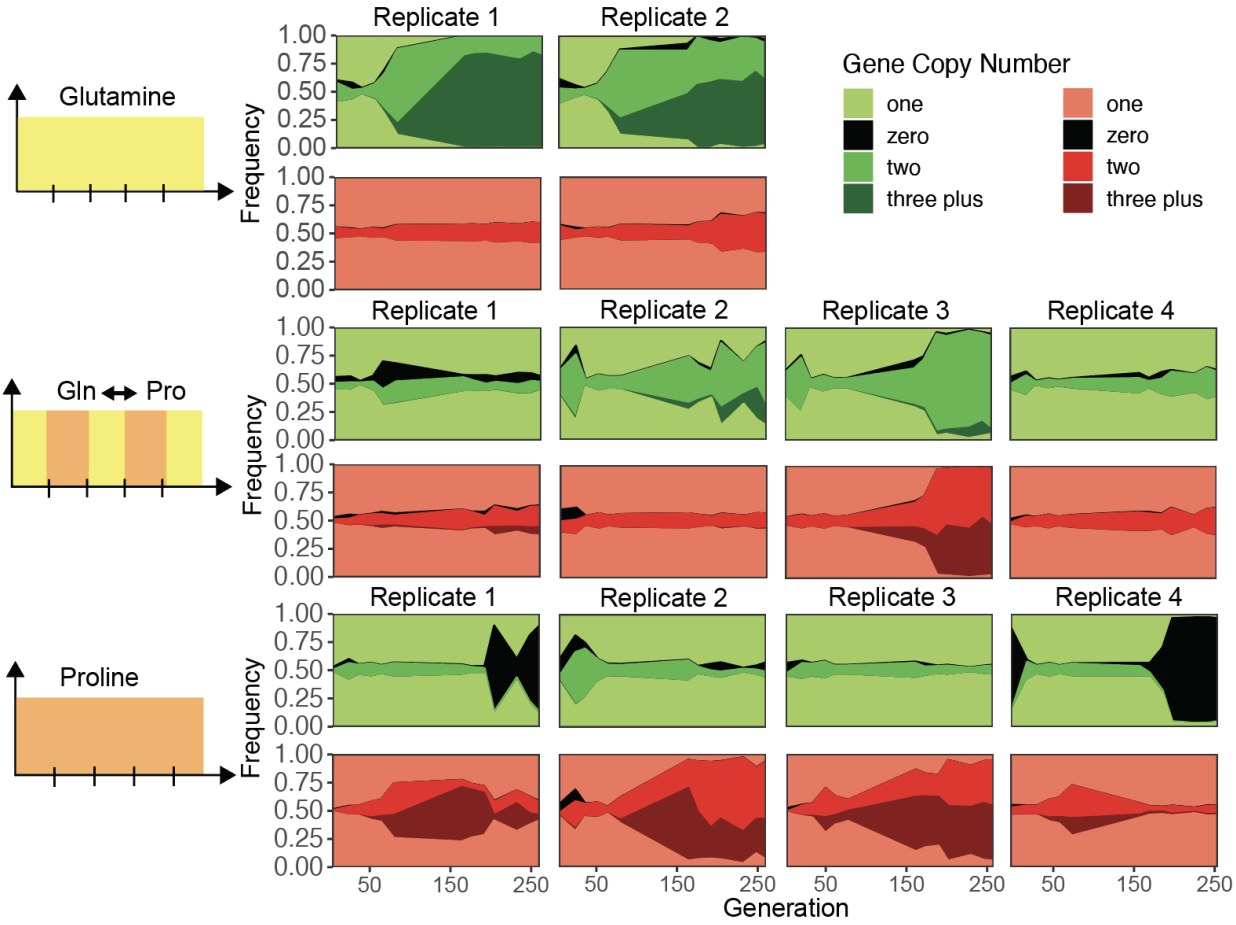


**Figure 3.3.3. Muller plots of copy number variants detected using the *GAP1-PUT4* CNV reporter**. The proportion of the populations containing zero, one, two and three plus copy number of either *GAP1* or *PUT4* is displayed. Green represents *GAP1* CNVs detected using the mCitrine CNV reporter and red represents mCherry signal reporting on *PUT4* CNVs.

### 3.3.4 Different gene copy numbers per cell are the drivers of adaptation at different temporal phases

Mean fluorescence gives a general summary of CNV frequency in a population, however, it is not informative about the diversity of classes of CNVs that contribute to the dynamics. To assess the contribution of individual CNV classes we used single-cell fluorescence profiles from flow cytometry analysis. We used the constructed fluorescent controls to define the fluorescent signal from zero, one, two, and three plus copies of a particular reporter. Muller plots were generated by assigning the parental strain to one copy of each allele tracked by either mCitrine or mCherry. In static glutamine-limitation, the initial rise in *GAP1* CNV frequency is a result of individuals with two copies of the allele, followed by a second wave of selection for individuals with three copies or more. While both conditions undergo complete sweeps of two or more *GAP1* CNVs, the decrease in total signal observed in replicate 2 relative to replicate 1, is explained by the smaller proportion of three plus *GAP1* CNVs. We also found that a subpopulation with two copies of *PUT4* arose in replicate 2 towards the end of the experiment suggesting that either *PUT4* CNVs have a previously unrecognized advantage in glutamine-limiting environments, or that *PUT4* duplications occur frequently and may have risen in frequency as a result of a hitch-hiking event.

### 3.3.5 A dual-color CNV reporter increases precision in detecting true CNV events

Changes in gene copy number in haploid yeast occur through mechanisms such as DNA polymerase slippage and nonhomologous end-joining. During experimental evolution, however, haploid yeast cells are known to self-diploidize to generate MATa/MATa and MATα/MATα genotypes (Venkataram et al. 2016) and in some instances spontaneous mating type switching results in a MATa/MATα genotype in otherwise clonal haploid populations (Harari, Ram, and Kupiec 2018). It is difficult to distinguish between an increase in fluorescence that is due to changes in copy number in haploid cells and an increase in fluorescence due to diploidization without performing additional assays that measure DNA content. Using two fluorescent gene reporters on different chromosomes results in a correlation between changes in mCitrine and mCherry fluorescence when both chromosomes are amplified as a result of the same mutational event such as the likely diploidization and the less likely whole genome duplication. In replicate 4 of the fluctuating condition between glutamine-limitation and ammonium-sulfate-limitation and in replicate 2 of the static ammonium-sulfate-limiting conditions, fluorescence in two loci mirror one another and are likely diploidization events (**Figure 3.3.2**).

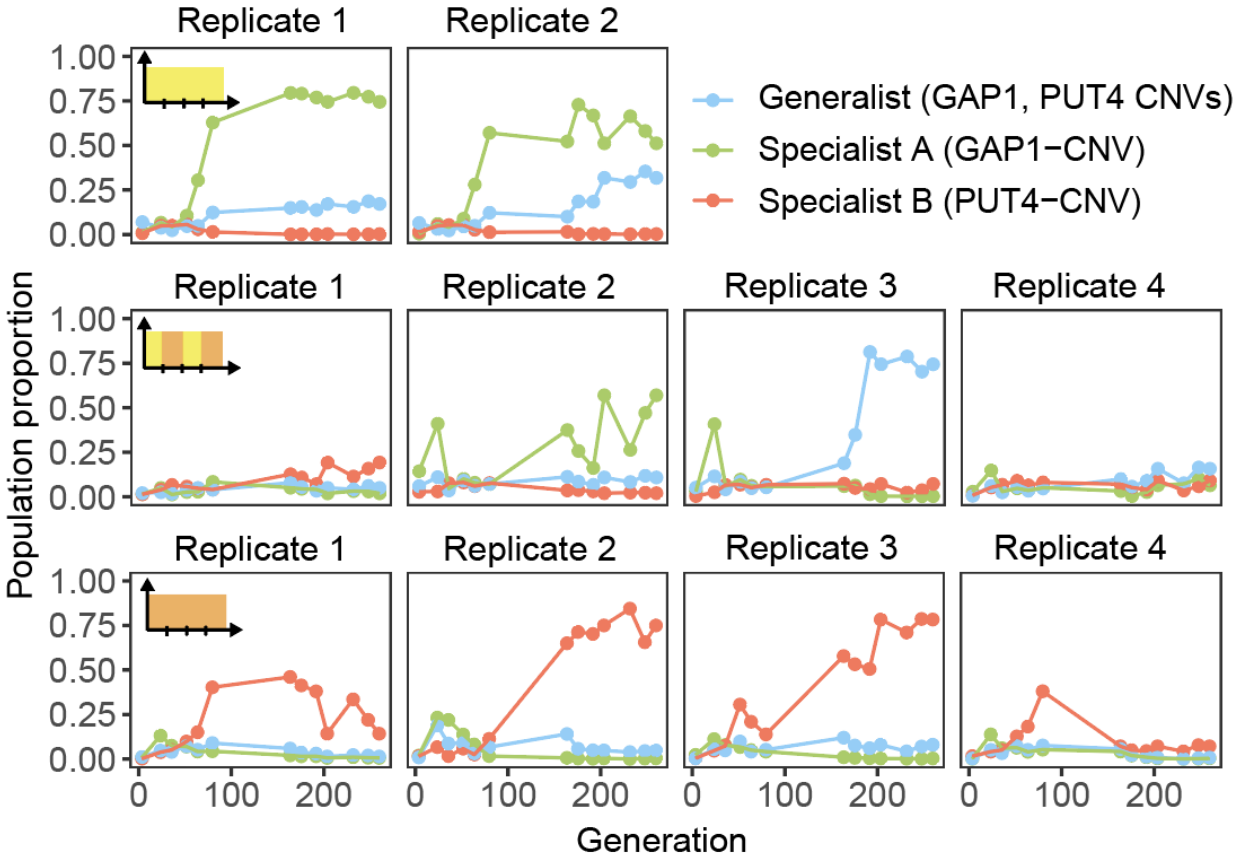


**Figure 3.3.4. The proportion of CNV specialists, and generalists in the *GAP1-PUT4* CNV reporter**. Specialists A are individuals with amplifications of *GAP1*, Specialists B are individuals with amplifications of *PUT4*, and Generalist defines individuals with amplifications of both *GAP1* and *PUT4*. The frequency of each strategy is shown across 250 generations with the top panel showing adaptation in glutamine-limitation, the bottom panel in proline-limitation and the middle panel in fluctuations between the two.

## 3.3.6 Fluctuating environments select for CNV-generalists

Understanding the change in copy number variation at two distinct loci is informative; however, understanding how CNVs within individual cells interact is informative about the evolutionary strategies co-opted in a population. Since CNVs can

readily form, they may be a rapid way in which an individual can switch back and forth between being a specialist and a generalist. We predicted that static conditions would select for specialists, i.e. individuals with increases in transporter gene copy number at a single locus that import the respective limiting nutrient. We found that in glutamine-limitation *GAP1*-CNV specialists rise to frequency as expected. However, we also find that CNV-generalists, in which individuals with increased copy number of both *GAP1* and *PUT4*, also increase in frequency in two replicate populations and comprise ~20-25% of individuals with *GAP1* CNVs at the final timepoints. This indicates that the initial rise in *GAP1* CNVs is determined by the rise in *GAP1* specialists, whereas, at later stages *GAP1-PUT4* generalists play a significant role in increasing CNV allele frequency. This result is unexpected as there is no known reason for which increased *PUT4* copy number would result in increased fitness during glutamine-limitation (**Figure 3.3.4**). Interestingly, in two other glutamine-limiting populations in which the *GAP1-MEP2* CNV reporter was used, a similar trend is observed despite the absence of *MEP2* CNVs in ammonium-sulfate-limitation (**Figure 3.7.4**).

In static proline-limiting conditions, PUT4-CNV specialists arise in all four replicate populations to frequencies between 30-75% of the population. Unlike the static glutamine-limiting condition, CNV generalists were not observed. The fluctuating environment shows the most heterogeneity in adaptive response. In replicate 1 the PUT4-CNV specialists reach the highest frequency at around 20%. In replicate 2 the GAP1-CNV specialist reaches the highest frequency at around 50%, however, this class of variation follows an oscillatory trajectory, consistent with specialist behavior in which

increases occur when glutamine is the sole nitrogen source and decreases occur when proline is the sole nitrogen source. Replicate 3 showed a sharp increase in frequency of the *GAP1-PUT4*-CNV generalist to a frequency of 80% suggesting that this strategy is beneficial under selection in fluctuating environments. Replicate 4 showed no visible increases in frequency in alleles for either *GAP1* or *PUT4* suggesting that adaptation has proceeded through a different mutational path, potentially SNVs.

These results suggest that static environments select for greater repeatability of outcome in comparison to fluctuating environments likely both in terms of the types of mutations that arise, as well as the selective pressure imposed despite the fluctuating environment being the same across four replicates and being periodic in nature.

## 3.4: Future and ongoing work

### 3.4.1 Fluctuating selection results in greater diversity

Several additional pieces of work must be undertaken in order to complete this study and these experiments are described below.

1) DNA has been sampled from the first experimental evolution, described in detail in this chapter, for whole population whole genome sequencing. I have performed sequencing for a fraction of the samples, however, additional sequencing must be performed to address questions about which mutations such as indels and SNVs and other structural variants exist in the population and how they vary over time. This will

also inform CNV dynamics as other mutations may be responsible for the changes in frequency and fitness of CNVs as evolution progresses.

2) While I was able to observe the dynamics of CNV frequency changes, it remained unclear whether the rise in frequency was due to strong selection or recurrence in mutation. It was also unclear whether reversion events were occuring, especially in fluctuating environments, in which one population underwent oscillations in CNV frequency. To address these questions unique barcodes were inserted by using the Levy barcoding system. Since we had previously observed negative fitness effects resulting from the placement of the barcode landing pad adjacent to the *DUR1,2* locus, for these experiments the landing pad was moved to the neutral locus, URA3, and 500,000 unique barcodes were inserted into the two experimental dual-CNV reporter strains. The long-term experimental evolution experiments we performed again but with the barcoded strains and DNA was sampled for barcode sequencing once a week. So far, I have identified that in time point zero we have barcode diversity that falls in the range of 60,000 - 200,000 barcodes per population, however, this experiment is ongoing and results will be completed in the near future. Additionally, in these new experiments I have added two additional static conditions with mixed nitrogen sources which are either proline + glutamine-limitation or ammonium-sulfate + glutamine-limitation. The goals of adding these conditions is to understand whether the temporal element of being exposed to one nitrogen source versus two nitrogen sources is the important factor in the dynamics and outcomes of evolution or whether static conditions that

deliver the same amount of nitrogen on average despite being static are responsible for the evolutionary dynamics that were observed.

3) The second long-term experimental evolution run performed which included barcodes will be important for additional analyses. I plan to map the trajectories of the unique barcodes in each population. I will investigate whether there is a condition-dependent effect on the maintenance of diversity, as was seen in Chapter 2, or not. Using barcoding data I will be able to quantify the fitness of each lineage in each population. Additionally, by combining fluorescence-activated cell sorting (FACS) with barcode sequencing I will be able to probe which specific class of CNVs is associated with which barcodes. For example by sorting cells that only have amplifications in *GAP1* versus cells that only have amplifications in *PUT4* versus cells that have both, I would be able to understand whether CNV reversions occur or not. If the same barcodes consistently make up one of the CNV classes through the 250 generations of the experiment, then this would support that reversions minimally contribute to the dynamics of CNV evolution.

4) Finally, I discuss the different classes of CNV individuals including *GAP1*-CNV specialists, *PUT4*-CNV specialists, and *GAP1*-*PUT4*-CNV generalists. While I hypothesize that having amplifications in both *GAP1* and *PUT4* results in genotypes that are beneficial under both glutamine-limitation and proline-limitation (i.e. generalists) this needs to be validated. I plan to isolate potential CNV-specialists and potential CNV-generalists and to perform fitness assays in three static conditions,

glutamine-limitation, proline-limitation, and glutamine + proline-limitation, and a condition fluctuating between glutamine-limitation and proline-limitation to understand the fitness effects of the different combinations of CNVs in cells.

## 3.5: Discussion

In this study we investigated the effect of environments that fluctuate periodically in nitrogen quality on the generation and selection of CNVs. We find that static conditions result in more repeatable dynamics with stronger selection in which CNVs rise in a frequency within the first 100 generations. These dynamics, however, were locus-dependent as they were observed for *GAP1* and *PUT4*, but not for *MEP2*.

We also found that fluctuating environments resulted in less predictable dynamics in which there was greater variation between population replicates. In some populations CNVs did not seem to be the main drivers of evolution, whereas in others CNV frequency oscillated periodically, likely matching the fluctuations in media type. And in one population we observed increases in CNVs at the two loci that were under fluctuating selection.

By classifying individuals with CNVs at two loci as generalists and individuals with CNVs at one locus as specialists, we observed that static conditions reliably selected for specialists, with the exception of glutamine-limitation in which it appeared that a proportion of the population was composed of generalists. Further characterization of these genotypes is required to determine whether these increases in

allele number are a result of diploidization or if they are true amplification at both loci. Finally we find that in fluctuating environments some populations have changes in CNV specialist frequency and in one population CNV generalists reach up 80% of the population suggesting that adaptation progresses through several different routes during environmental fluctuations which may be determined by which mutations occur first early during evolution.

This study sheds light on the important effect of CNV formation during evolution in temporarily fluctuating conditions and suggests that further studies are required to understand the contribution of other types of mutation in determining the different possible routes in which evolution can progress.

## 3.6: Methods

### 3.6.1 Media and growth conditions

Nitrogen-limiting media (glutamine, proline, and ammonium-sulfate) contained 800 μM nitrogen regardless of molecular form and 1 g/L $CaCl_2$-$2H_2O$, 1 g/L of NaCl, 5 g/L of $MgSO_4$-$7H_2O$, 10 g/L $KH_2PO_4$, 2% glucose and trace metals and vitamins as previously described (Hong and Gresham 2014). Media with two nitrogen sources contained 400uM nitrogen of each source to a total of 800uM nitrogen.

We inoculated the *dual-fluorescent* CNV reporter strains into 20-mL ministat vessels (A. W. Miller et al. 2013). For static conditions either glutamine-, ammonium sulfate-, or proline limited media was continuously pumped into vessels. For fluctuating conditions in which we alternated between two media types every 48 hours, we

manually switched media type by connecting two media carboys to each ministat vessel and using plastic clamps to close one inlet and open the other. Control populations containing either one or two copies of the CNV reporter replacing neutral loci (*HO* and *YLR122/23C*) were also inoculated in ministat vessels for each media condition. Ministats were maintained at 30°C in aerobic conditions and diluted at a rate of 0.12 $hour^{-1}$ (corresponding to a population doubling time of 5.8 hours). Steady-state populations of $3 \times 10^7$ cells were maintained in continuous mode for 250 generations (60 days). Every _ generations, we archived 2-mL population samples at −80°C in 15% glycerol, and 2mL in storage solution (0.9M sorbitol, 0.1M EDTA, 0.1M Tris).

## 3.6.2 Strain construction

We used FY4, haploid and diploid derivatives of the reference strain S288c, for all experiments. S1 Table is a comprehensive list of strains constructed and used in this study. To generate dual-fluorescent CNV-reporter strains, we used the CNV-reporter strain generated in (Lauer et al. 2018) with the constitutively expressed *mCitrine* gene marked by the KanMX G418-resistance cassette (*TEFpr*::*KanMX*::*TEFterm*) inserted adjacent to the *GAP1* locus as our base strain. We constructed a plasmid using gibson assembly where the cassettes were generated by PCR amplification of the entire *mCitrine* plasmid with the exclusion of the mCitrine ORF and the PCR amplification of the ORF of mCherry from a plasmid from___. We performed high-efficiency yeast transformation (Gietz and Schiestl 2007) with the *mCherry* gene under control of the constitutively expressed *ACT1* promoter (*ACT1pr*::*mCherry*::*ADH1term*) and marked by the HygR G418-resistance cassette (*TEFpr*::*HygR*::*TEFterm*) and inserted it either

adjacent to *PUT4 (*integration coordinates, Chromosome XIII: 000–000*)*, or *MEP2 (*integration coordinates, Chromosome XIII: 000–000*)*. For 1-copy control strains, either the *mCitrine* or *mCherry* reporter was integrated at one of two neutral loci: either replacing *HO* (*YDL227C*) on Chromosome IV or replacing the dubious ORF, *YLR122/23C*, on Chromosome XII. To generate either 2-copy *mCitrine* or *mCherry* haploid controls, we mated the 1-copy controls, sporulated and dissected the resulting diploids. PCR and Sanger sequencing were used to confirm integration of the CNV reporters at each location.

For lineage tracking, we constructed a strain containing a landing pad and the *dual-fluorescence* CNV reporters by reconstructing the landing pad (LP) plasmid by replacing the kanMX marker with natMX. We PCR amplified the LP fragment using primers with 40bp homology to the upstream and downstream region of URA3. A high efficiency transformation was performed to integrate the cassette replacing the URA3 gene. Transformants were plated on natMX plates to select for positive transformants. Transformants were confirmed by replica plating on 5-FOA plates and through colony PCR of the 5' and 3' junctions, and subsequent sanger sequencing of the products. Whole genome sequencing of these strains was also performed. We introduced a library of random barcodes (Levy et al. 2015) by transformation and selection in liquid agar. mCitrine controls were used from Lauer et al. 2018, and mCherry controls were designed using an mCherry-hyg strain by making primers with 40bp homology to amplify the DNA fragment. A high efficiency yeast transformation was performed.

## Acknowledgements

We thank previous and present members of the Gresham, Vogel and Ghedin labs lab for valuable discussion and comments on the manuscript. We thank the Kerry Geiler-Semarotte lab for primer sequences used for 96-well multiplexing of barcoded libraries. We thank Charles Miller for tetrad dissections and Erika Levine for early CNV qPCR experiments and analyses. We thank the NYU Genomics Core facility for sequencing services and specifically Hana Husic for flow cytometry services. This work was supported by

## 3.7: Supplementary figures

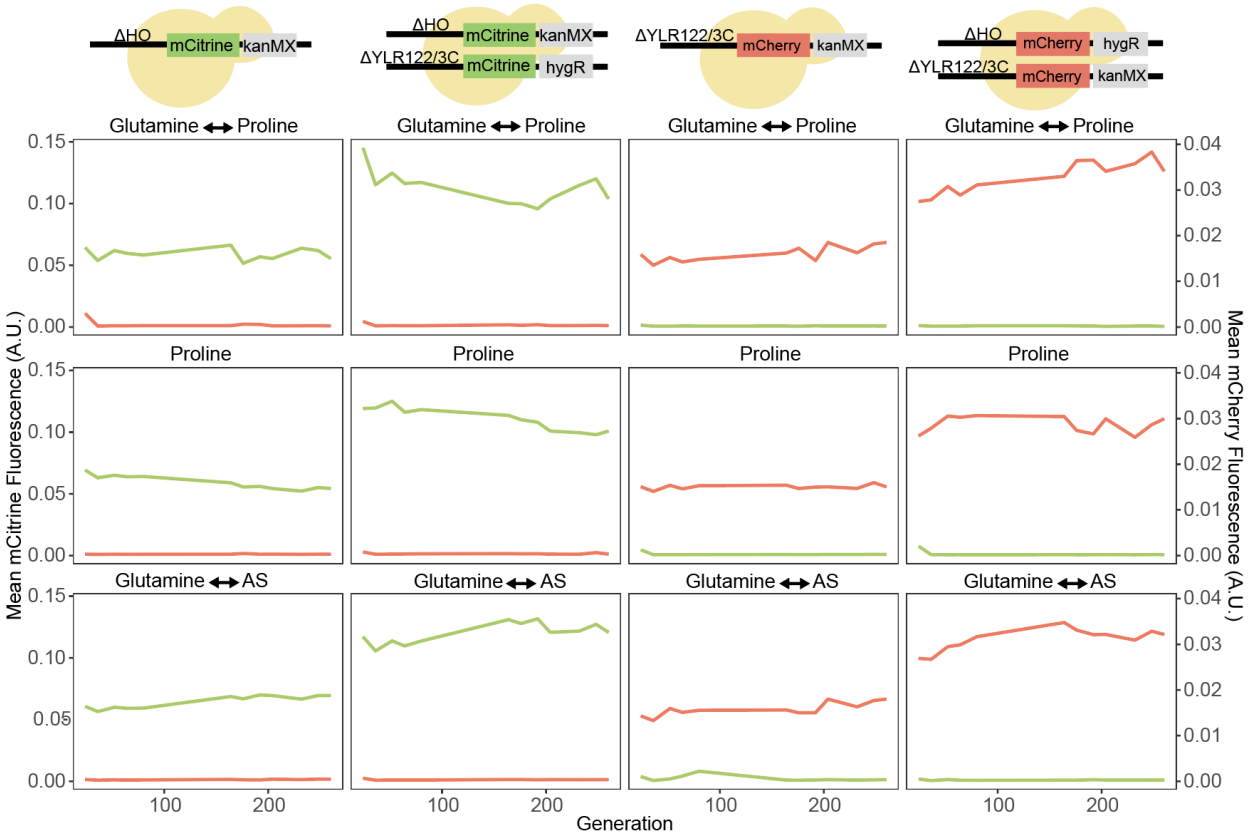


**Figure 3.7.1. Control strains experimentally evolved in three distinct conditions for 250 generations**. Control strains include either 1 copy mCitrine, 1 copy mCherry, 2 copies mCitrine, or 2 copies of mCherry inserted at neutral loci. The red line represents mCherry fluorescence levels, and the green line represents mCitrine fluorescence levels.

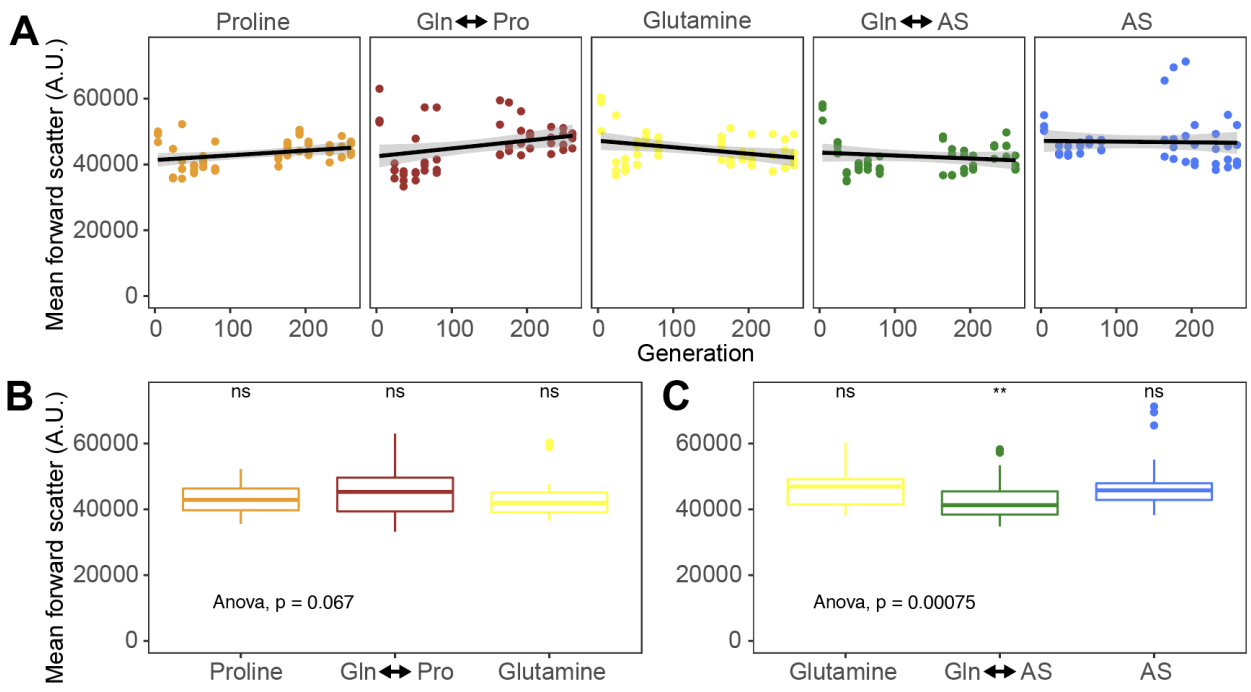


**Figure 3.7.2. Forward scatter, a proxy measurement for cell size, shows minimal change across conditions**. (A) Mean forward scatter across experimental evolution time is shown for each condition which is made up for replicate populations. (B) Boxplots of mean forward scatter for static proline-limitation (Pro) and glutamine-limitation (Gln) and fluctuating between proline-limitation and glutamine limitation with the anova test performed pairwise for all three conditions, n = 52. (C) Boxplots of mean forward scatter for static ammonium-sulfate-limitation (AS) and glutamine-limitation (Gln) and fluctuating between proline-limitation and glutamine-limitation with the anova test performed pairwise for all three conditions and a t-test to determine significance, n = 52.

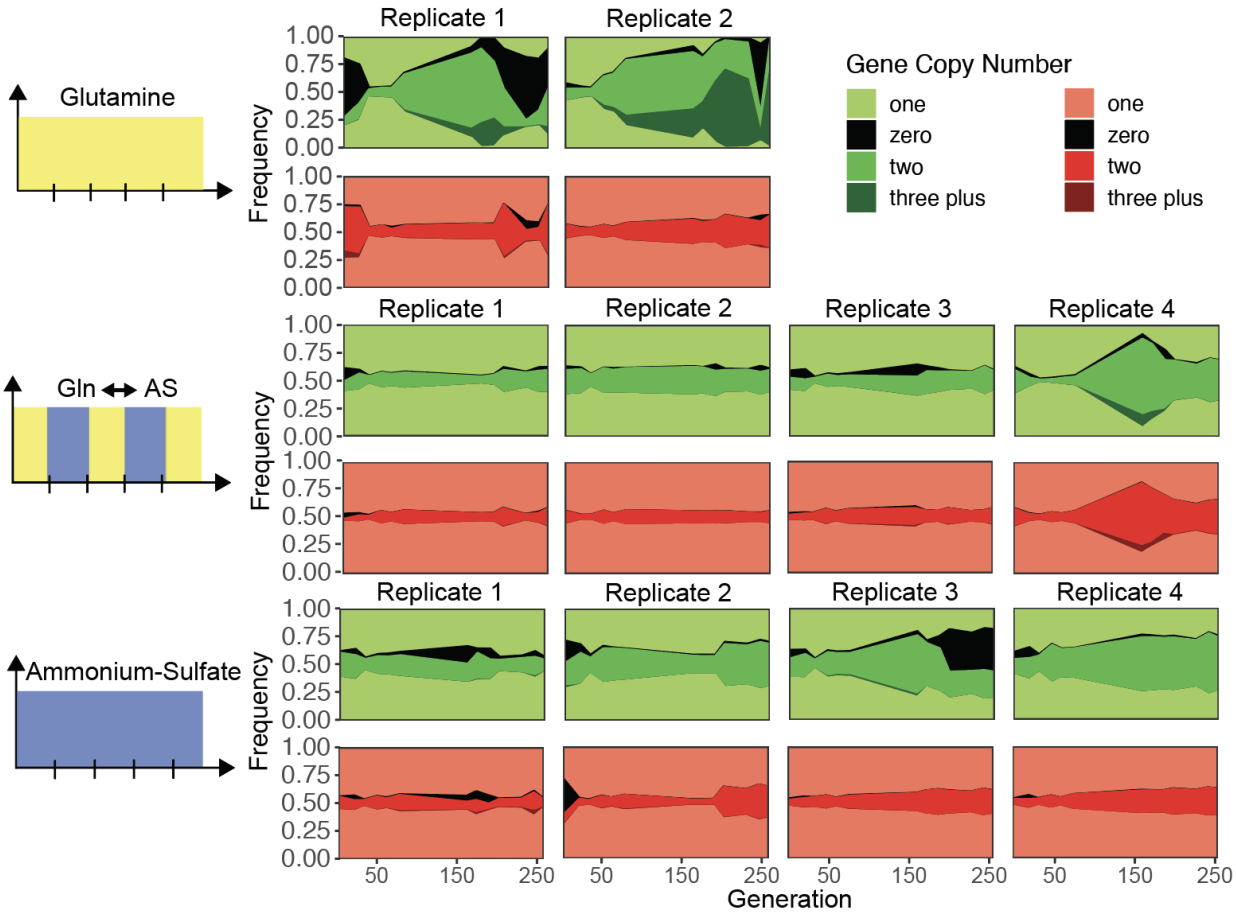


**Figure 3.7.3. Muller plots of copy number mutations in the *GAP1-MEP2* CNV reporter**. The proportion of zero, one, two and three plus copy number of either *GAP1* or *MEP2* is visualized in muller plots. The green muller plots represent mCitrine and by proxy *GAP1* and the red muller plots represent mCherry and by proxy *MEP2*.

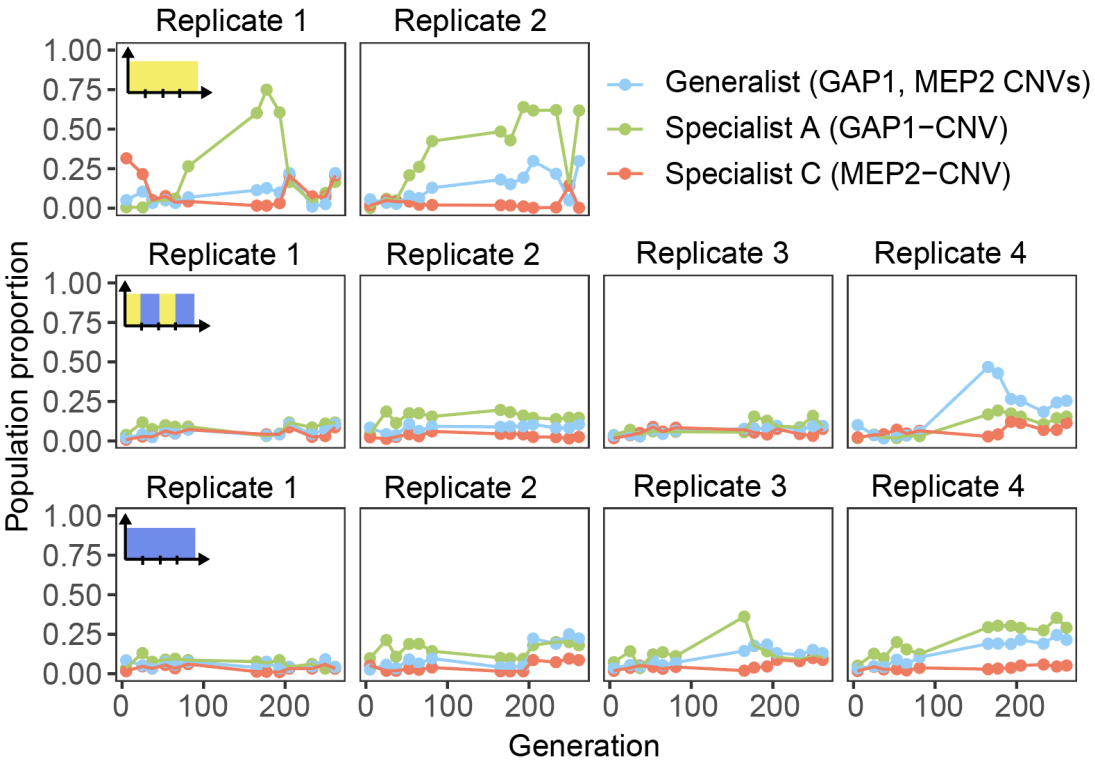


**Figure 3.7.4. Plots showing proportion of CNV specialists, and generalists in the *GAP1-MEP2* CNV reporter.** Specialists A are individuals with amplifications of *GAP1*, Specialists B are individuals with amplifications of *MEP2*, and Generalist defines individuals with amplifications of both *GAP1* and *MEP2*. The frequency of each strategy is shown across 250 generations with the top panel showing adaptation in glutamine-limitation, the bottom panel in ammonium-sulfate-limitation and the middle panel in fluctuations between the two.

# Chapter 4: Constructing a new lineage tracking barcode system

This chapter is based on work performed with Dr. Chris Jackson.

## 4.1: Abstract

Lineage tracking is a useful method for the study of adaptive evolutionSynthetic barcodes are random DNA sequences that are heritable and when inserted into the genome of individual cells in isogenic populations, allow the identification of descendents of particular ancestral cells for hundreds of generations. Recently Levy et al. developed a synthetic DNA barcode system for use in *Saccharomyces cerevisiae* and it has been widely used to address important evolutionary questions about the dynamics of clonal interference and genetic diversity in evolving microbial populations. Although the Levy system has been effective in pioneering the use of barcoding in microbial populations, we have observed that it causes a fitness reduction as a result of positional effects and genetic load. Here we present the development of a novel CRISPR-based barcoding system that is simple to use and minimizes negative fitness effects.

## 4.2: Introduction

Tracking how mutations arise and are selected in populations poses several technical challenges that have hindered the study of adaptive evolution. The dynamics of evolution are a product of both the mutation rate and selection coefficients, making it difficult to tease apart the respective contributions of each process. Historically, the

dynamics of evolution have been tracked through the measurement of phenotypic and morphological traits. However, traits that can be traced to multiple genetic loci have made these measurements susceptible to erroneous conclusions. In the age of genomics in which high throughput DNA sequencing is used to identify specific mutations, our ability to tease apart genetic loci in polygenic traits has significantly improved. However, tracking the frequency of a mutation does not distinguish between mutation recurrence and selection.

DNA barcodes are stretches of DNA sequence that can be used as unique cellular identifiers and have been implemented in a variety of experimental designs. DNA barcodes are at the foundation of how phylogenetic relationships are resolved. Carl Woese was the first to apply the use of DNA “barcodes” in conjunction with DNA sequencing, when he constructed the tree of life using the endogenous variable region of rDNA, a DNA sequence that is ubiquitous across all domains of life. In recent decades the use of synthetic DNA barcodes, which are artificially engineered DNA sequences, has exploded in a variety of applications including, genetic screens, multiplexing of sequencing samples (Lyons et al. 2017) and lineage tracking (Lyons et al. 2017; Levy et al. 2015).

Levy et al. developed a method in which individual cells in a population are uniquely tagged with synthetic DNA barcodes that are integrated into the genome. When the barcoded population is propagated over hundreds of generations under a selective pressure, each lineage can be identified and tracked as populations adapt.

Tracking lineages over generations opens the door to studying the dynamics of evolution with unprecedented precision. Combining barcode sequencing with whole-genome DNA sequencing distinguishes between whether mutations have arisen recurrently or whether a single sweep is responsible for the increase in frequency of a particular mutation. DNA barcodes also allow for the study of how diversity in a population changes over time (Blundell et al. 2017) and the performance of pooled fitness assays of evolved individuals instead of classic pair-wise fitness assays (Blundell et al. 2017; Venkataram et al. 2016). The use of DNA barcodes during experimental evolution permits the high-resolution visualization of the process of adaptation and the tackling of several age-old questions about adaptive evolution processes.

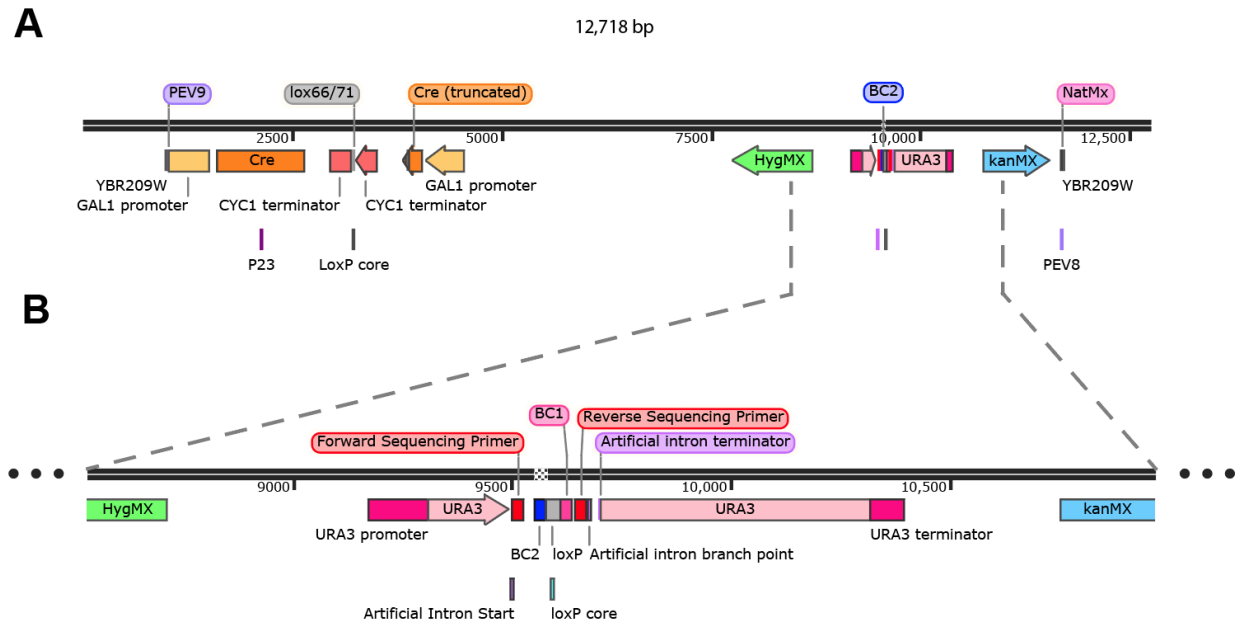


**Figure 4.2.1 Levy et al's barcoding system.** (A) 12,718 additional base pairs are inserted and retained in the genome for successful barcoding. Features are annotated including BC2 which is the site of the 26bp unique DNA barcode. (B) The region that contains the DNA barcode required for lineage tracking.

Levy et al.'s barcoding system works using a two-step transformation in which a "landing pad", which includes the Cre recombinase and half of URA3, is first integrated into the genome in a region adjacent to DUR1,2. A plasmid library with a backbone including the second half of URA3 and an integrated barcode with high complexity is then transformed into the *Saccharomyces cerevisiae* strain. The galactose-inducible Cre recombinase, a bacteriophage topoisomerase that catalyzes the site-specific recombination of DNA between loxP sites, is activated by the addition of galactose. Cre mediates the integration of the unique barcodes demarcated the loxP site in the landing pad. While the DNA barcode is only 26bp long resulting in minimal alterations to the genome, the entire cassette is maintained in the genome, representing over 12kb of additional DNA (**Figure 4.2.1**). The location of such a large construct can have significant positional effects. Indeed, we have found that yeast cells barcoded with this system are unable to grow in media with urea as the sole nitrogen source likely a result of a perturbation to the adjacent Urea metabolism genes including the Urea amidolyase encoded by Dur1,2, which degrades urea to CO2 and NH3.

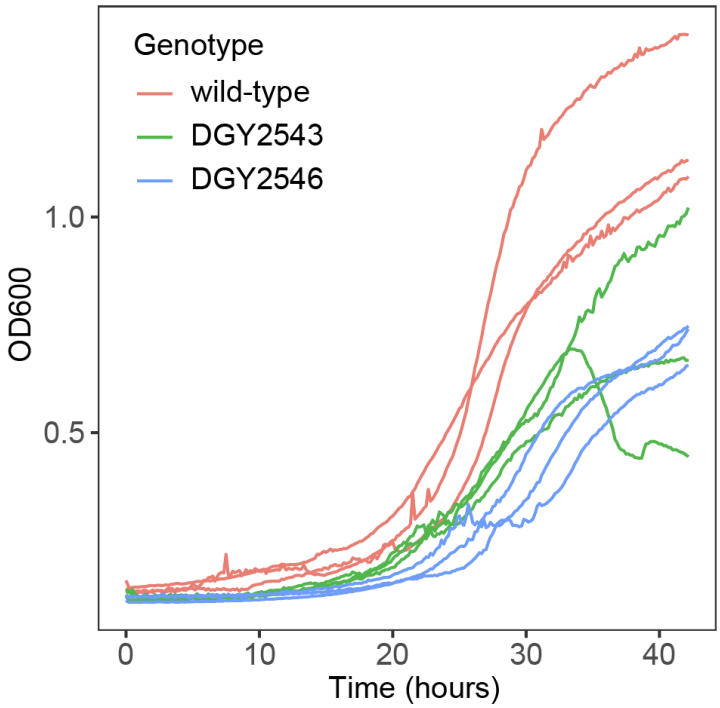


**Figure 4.2.2 Fitness assays of two genotypes barcoded with the Levy barcoding system.** Growth curve assays were performed in triplicate for three different strains in batch proline-limiting media.

We have attempted to alleviate this issue by moving the Levy barcoding system to the neutral URA3 locus (**Chapter 3**). We constructed two strains DGY2543 and DGY2546 (described in detail in **Chapter 3**), each with a different genetic background, which were barcoded with the Levy system at the URA3 locus. Doing so showed a slight but noticeable defect during growth in proline-limitation in comparison to the wild-type strain further underscoring the importance of creating a less burdensome barcoding system. This effect was evident both in showing an increased lag phase and a lower growth rate during exponential phase (**Figure 4.2.2**). While barcoding is an important tool for studying the dynamics of evolution, there remain multiple issues that must be addressed to ensure the creation of a versatile and non-deleterious barcoding system.

In this chapter, I describe construction of a novel barcoding system for lineage tracking during experimental evolution, with the goal of creating a system that minimizes genetic load, with minimal positional-effects, and is simple to construct and use.

## 4.3: Results

One of the issues with inserting synthetic DNA barcodes into genomes is that the ability to target them efficiently into a specific location requires the insertion of other regulatory genetic material. In the Levy barcode system, a ‘landing pad’ is inserted through homologous recombination adjacent to DUR1,2, which is used to target the insertion of the plasmid barcode library to that location by Cre-lox. The edited locus after the completion of transformation, includes many regulatory elements for the process, including a full Cre ORF, a truncated Cre ORF, and numerous unessential drug markers, one of which is the *E. coli* drug marker AmpR. The location and size of additional DNA are deleterious in certain conditions. To bypass this issue we designed a system that results in a cleanly barcoded region at a neutral locus in the genome.

### 4.3.1 Overview of new barcode lineage tracking system

CRISPR/Cas9 has been widely used for targeted genome editing because of its high efficiency. CRISPR systems are composed of two components: 1) a guide RNA (gRNA) sequence that is complementary to the region that is the target of editing and 2) CAS9, the protein that cleaves the DNA that the gRNA targets enabling the insertion/deletion of DNA. Transfection of CAS9 is either stable or transient and is

frequently delivered by a vector (H. Wang, La Russa, and Qi 2016). It is attractive to have a stable system when cell lines must be propagated, however, in instances where CAS9 will be used once to edit the genome, it is preferable to delete or suppress its expression afterwards as off-target effects are known to occur in the presence of constitutively expressed CAS9.

We designed a barcoding system in which CAS9 is galactose-inducible, to minimize off-target effects when it is inactivated. CAS9 is integrated into the genome at the neutral locus YPRCdelta15, which is an inactive retrotransposon that has lost functionality. A “suicide casette”, containing the gRNA sequence adjacent to its target sequence, an exogenous sequence from *Bacillus subtlis*, is inserted replacing the URA3 locus. CAS9 mediates the insertion of a cassette from a high-complexity plasmid library with a unique barcode adjacent to the wild type URA3 sequence replacing the original “suicide casette” restoring wild type function of URA3 and resulting in a DNA barcode that is ~20bps in size with no additional surrounding sequence coding for regulatory machinery with the exception of the inducible CAS9 and nourseothricin resistance drug marker (NatR) at YPRCdelta15 which is 5,642 bps long, less than half the size of the levy barcoding system (**Figure 4.1.1**).

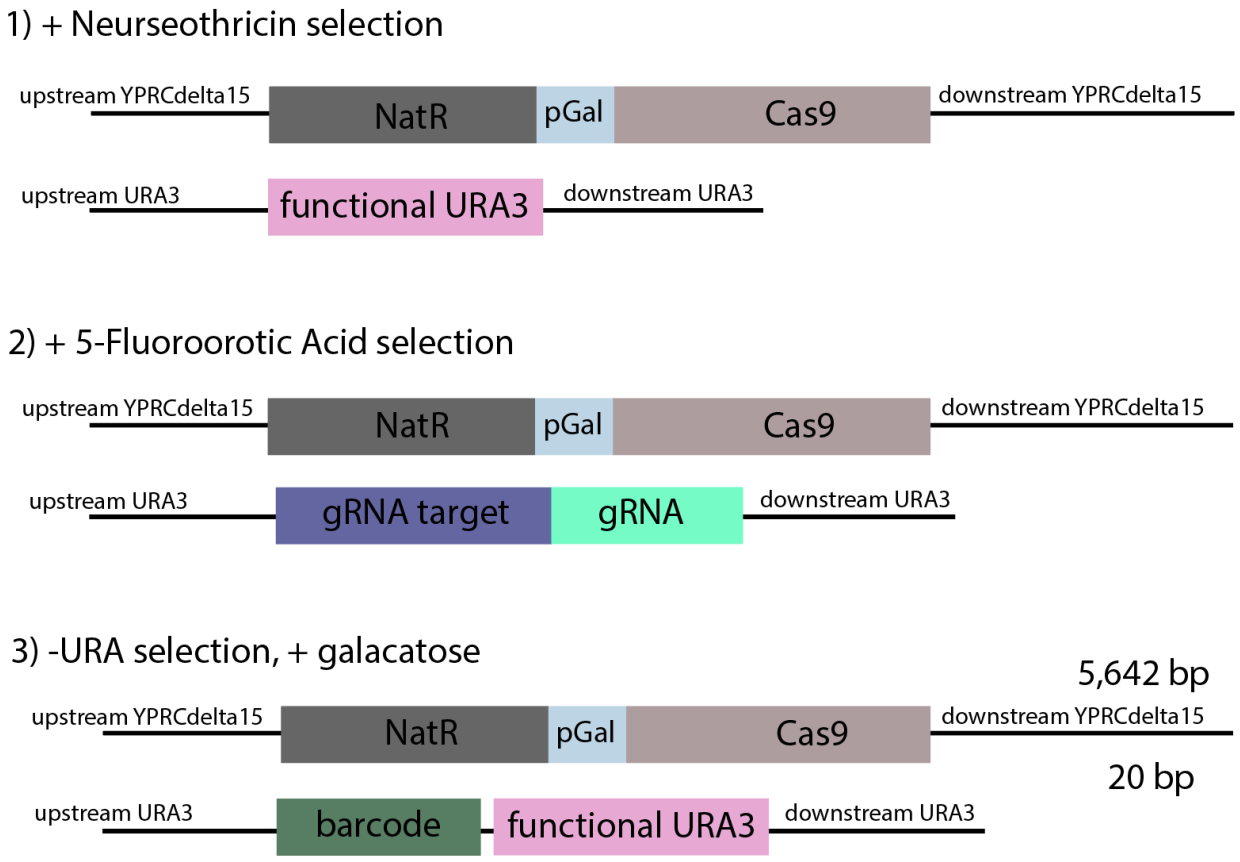


**Figure 4.3.1 Lineage barcoding design overview.** (1) + Neurseothricin selection for the insertion of NatR-Cas9 cassette at YPRCdelta15 neutral locus. (2) + 5-FOA selection for the gRNA suicide casette at the URA3 locus. (3) - URA3 selection with galactose enables Cas9 to excise the gRNA suicide casette and replace with the unique lineage tracking barcode plasmid library.

## 4.3.2 Constructing a Gal-inducible Cas9 yeast strain

We constructed a Gal-inducible CAS9 plasmid by Gibson assembly of the native yeast GAL1,10 promoter, the CAS9 coding sequence, and a constitutively expressed NatMX cassette (**Figure 4.3.2A**). We performed selection on media with Nourseothricin and identified five potential transformants. We performed a confirmation PCR using a three-primer design (**Figure 4.3.2B**) and isolated the strain DGY2280. We sought to validate whether CAS9 expression was functioning as expected and as such we grew

DGY2280 in galactose and glucose as the control and performed a western blot with an anti-CAS9 antibody. We found that protein expression of CAS9 was minimal at time zero in both conditions but in galactose expression was strongly expressed after 8 hours confirming that CAS9 is inactive in the absence of galactose and is activated in the presence of galactose. Cas9 is around 160 kDa in size and while our gel is not able to fully resolve this size, we do see the band between 150-250 kDa as expected in a gel with low resolution. There was a band that is believed to be noise in the western blot which ran at around 60kDa and was present in both conditions across all timepoints (**Figure 4.3.2C**). We found no noticeable growth effects as a result of the genetic manipulation (**Figure 4.3.2D**).

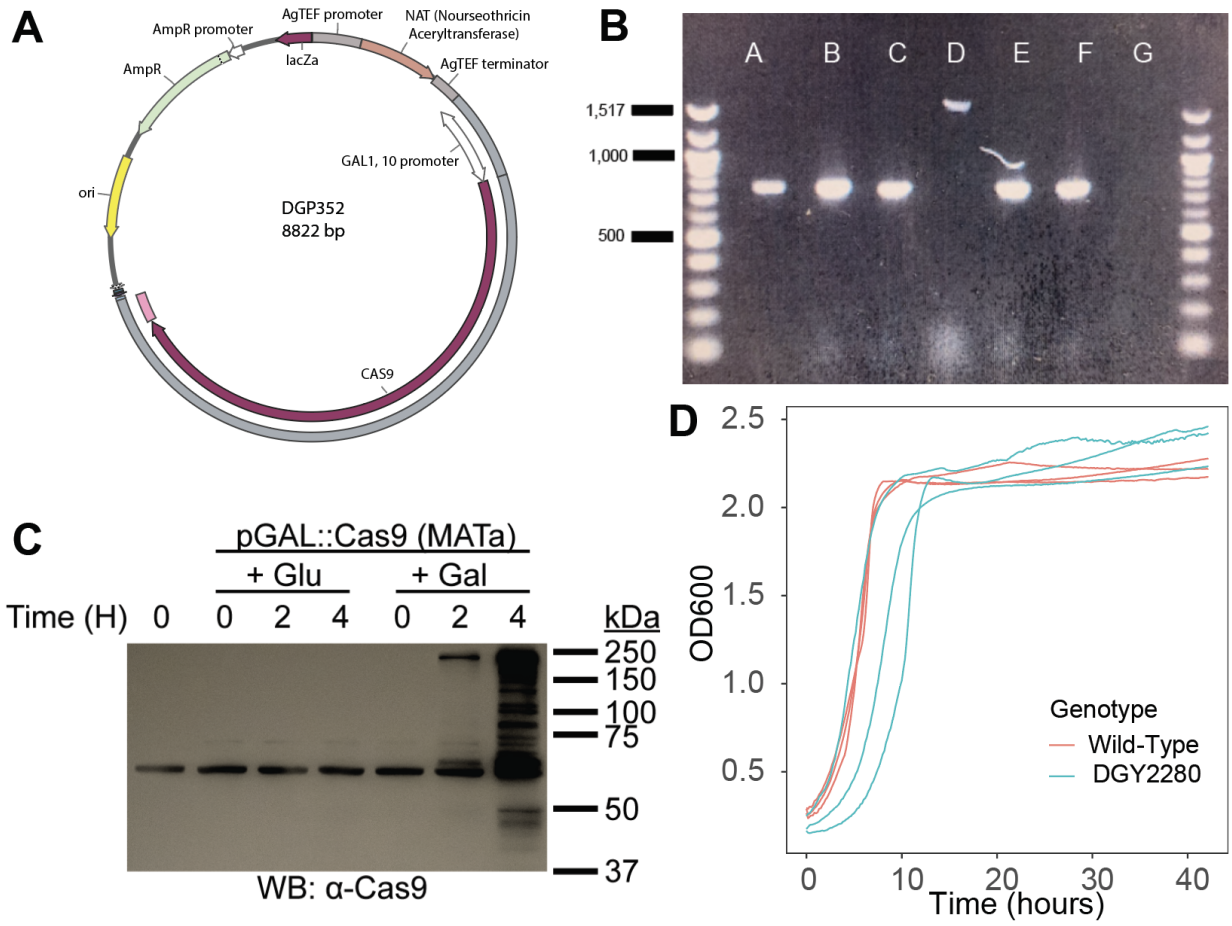


**Figure 4.3.2 Insertion of galactose-inducible Cas9 at a neutral locus.** (A) Assembled plasmid with the NatMX cassette, gal promoter, and Cas9 coding sequence. (B) PCR gel of potential transformants with the Cas9-NatMX cassette inserted at the YPRCdelta15 locus in lanes A, B, C, E and F. Lane D is a negative control where the Cas9-NatMX cassette was not inserted and lane G is a negative control with no DNA template. (C) Anti-Cas9 western blot of a time series where DGY2280 is grown on either glucose or galactose. (D) Growth curves of wildtype yeast and DGY2280 in YPD + 2% glucose.

## 4.3.3 Constructing gRNA plasmids

To construct the gRNA “suicide cassette” plasmid we assembled the base plasmid, DGP353, with GFP flanked by two different exogenous sequences from *Bacillus subtilis* as orthologous gRNA targets that are unique and absent from the yeast genome to minimize accidental DNA cleavage in other locations in the genome. To

ensure that the gRNA efficiently targets it's complementary sequence, we designed three different gRNAs targeting unique regions in the *B. subtilis* sequence. In addition, we designed a negative control comprising a gRNA that is not complementary to the *B. subtilis* sequence or any other sequence in the yeast genome. To ensure that the gRNA does not have off-target effects, growth rate will also be compared to DGY2280 the strain without a gRNA. The four different double stranded gRNAs were each independently inserted into DGP353 to make four plasmids, in which successful insertion replaced GFP and resulted in non-fluorescent *E. coli* (**Figure 4.3.3B**). With these constructions I have laid the foundation for the creation of a new barcoding system, however, several steps remain to be performed to complete and test this system.

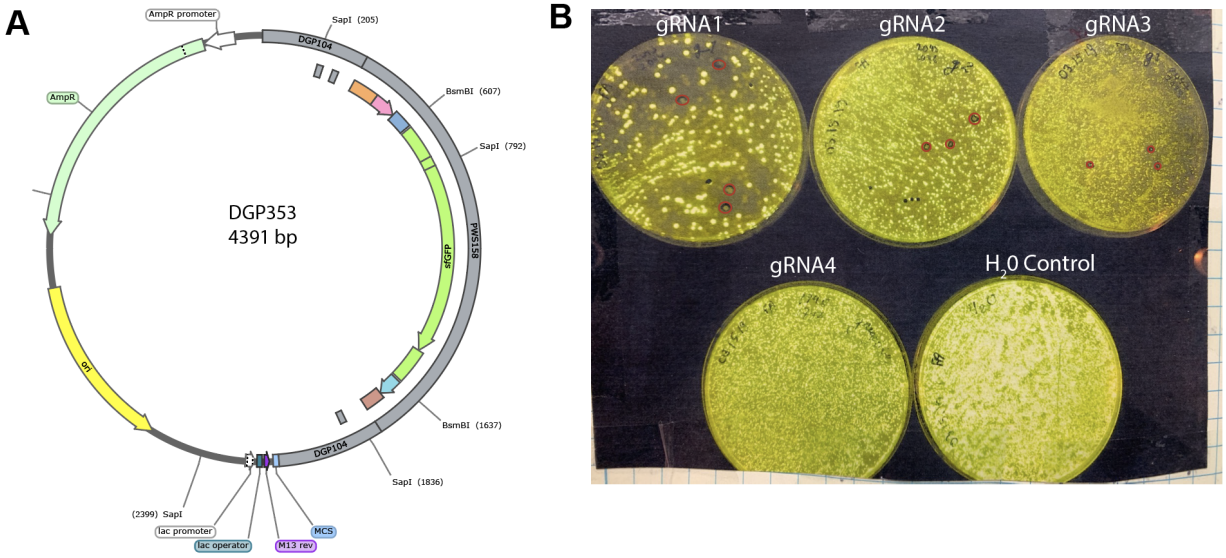


**Figure 4.3.3 Golden gate performed to create the gRNA plasmid.** (A) DGP353 is the plasmid constructed with the base suicide casette which includes GFP before the gRNA sequence is inserted replacing GFP. (B) Golden gate assembly is used to replace GFP in DGP353 with four different potential gRNAs and a water control. Successful replacements, circled with red, result in colonies that do not glow under blue light.

## 4.4 Future work: Creation of a high-complexity barcode plasmid library

In the future, the PCR amplification of each of the “suicide cassettes” must be performed and followed by the transformation of each cassette independently into DGY2280 using 5-FOA selection to select for successful integration (**Figure 4.3.1**). This would result in three potential barcoding strains and a control strain.

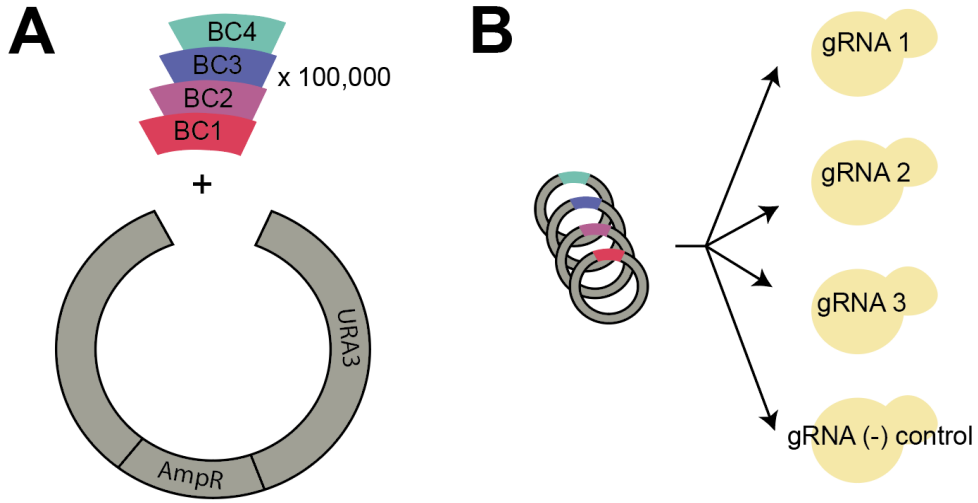


**Figure 4.4.1 Completion of the barcode system.** (A) The construction of the plasmid library requires a backbone with AmpR for ampicillin selection E. coli and wildtype URA3 for selection during transformation into yeast. 100,000 DNA barcode oligos are inserted into the backbone through restriction digest and ligation (B) The plasmid library is the transformed into four different yeast strains with either gRNA 1, gRNA 2, gRNA 3, and the negative gRNA with no complementation in the genome.

A plasmid library will be constructed by synthetically ordering one hundred thousand unique degenerate oligonucleotides that are 20 base pairs long. These oligonucleotides will be inserted into a plasmid upstream of the wildtype URA3 gene. A golden gate cloning reaction using a Type IIs restriction enzyme in a cyclic reaction with T4 DNA polymerase will insert the degenerate library into the URA3 replacement cassette to create a high complexity barcode library. This library will be transformed into *E. coli* and grown out in a 1L soft agar selection media, which allows bacterial colonies to grow in three dimensions without clonal interference (**Figure 4.4.1A**). The plasmid

library will be extracted with a standard maxiprep and transformed into the four constructed yeast strains in the presence of galactose to activate the transcription of Cas9. Each yeast strain has one of the four gRNA sequences, three that are expected to have a complementary sequence as part of the “suicide casette” that is transformed into the yeast, and a fourth that is expected to not have a genomic complementary sequence and there would act as negative control (**Figure 4.4.1B**). gRNAs might have lowered efficiency of reaction despite having a complementary sequence which is why it is important to test three different sequences. The control is important to set a baseline transformation efficiency that is not dependent on the CRISPR/Cas9 system and complementarity. Transformation efficiency will be assessed by observing the percentage of transformed cells that result in colony forming units on the URA- selective plates. Strains will then be tested for barcode complexity by performing amplicon sequencing on the Illumina NovaSeq6000 platform and counting unique barcodes by using the Bartender algorithm. The strain with the highest transformation efficiency and DNA barcode complexity can then be used for lineage tracking experiments. Once the these initial experiments are performed and it is known which gRNA plasmid results in the greatest barcode complexity, future experiments in which lineage tracking may be performed in strains with different genetic backgrounds, the process would only require the three-step transformation using the single efficient “suicide casette”.

## 4.5: Discussion

We have laid the foundation for the creation of a new barcode lineage tracking system that is minimally invasive and in theory can be easily constructed in any yeast

strain. By combining the Gal-inducible Cas9 with a 3-step transformation it is possible to establish this novel barcoding system.

## 4.6: Conclusion

Lineage tracking is a powerful method to tackle questions about the laws of evolution. The present theoretical framework surrounding evolution in fluctuating environments is based on a combination of decades of hypotheses, indirect observations of the outcomes of evolution, and mathematical modelling, with a dearth of real examples of evolution observed in real-time. With the rapid technological advances of synthetic biology and next generation sequencing it has become possible to directly track the dynamics of the molecular bases of evolution with unprecedented precision. The use of synthetic DNA barcodes allows researchers to have experimental models of evolution in changing environments to make the leap from purefly theoretical hypotheses of how adaptation occurs to an understanding rooted in empirical evidence. The use of DNA barcodes is not only useful in experimental models but is also imperative in studying evolution in natural populations such as in models of the gut microbiome, which would have medical implications by informing the design of probiotics and antibiotics. DNA barcodes are also useful when used in bacterial systems that are known to undergo both vertical gene transfer through replication, and horizontal gene transfer through conjugation and other within generation methods of sharing DNA. The dynamics of horizontal gene transfer have been especially challenging to study since they occur over very rapid timescales and can be elusive to detection through

DNA sequencing alone. The useful applications of synthetic DNA barcodes for lineage tracking highlight the importance of the development of an easy to use system with minimal fitness consequences.

## 4.7: Methods

### 4.7.1 Construction of plasmids

Three plasmids were constructed to DGP352, DGP353, DGP354 (Table 4.1). All plasmids were cloned using Gibson assembly in a pUC19 backbone and constructions were confirmed using sanger sequencing. DGP352 was constructed by combining GAL promoter driving Cas9 expression and NatR from the same bidirectional promoter. NAT is from DGP328. Cas9 from DGP290. Gal1/10pr from genomic DNA. NAT terminator is tPRM9 from yeast genomic DNA. DGP353 was constructed by combining a sequence from *Bacillus subtilis* taken from DGP104. The *B. subtilis* sequence is interrupted by a sequence from DGP290 that contains the promoter elements required for gRNA expression and a GFP gene. The GFP can be excised using BsmBI to insert specific gRNAs. DGP354, DGP355, DGP356, DGP357 were constructed by using DGP353 as the base plasmid and using golden gate assembly to replace GFP with the respective guide RNA. gRNAs were prepared by annealing complementary primers by heating in a heat block to 95C for 5 minutes and then removing the heat block to allow for slow cooling for 45 minutes. Four gRNAs were designed where three were positive tests that are targeted to different locations in the *B. subtilis* sequences and one gRNA is the negative control constructed to target no sequence.

**Table 4.6.1. Plasmids used in this chapter**

| Plasmid Number | Plasmid Name |
|---|---|
| DGP352 | pUC19-GAL1-Cas9 |
| DGP353 | pUC19-gRNA-SUICIDE |
| DGP354 | pUC19-gRNA1 |
| DGP355 | pUC19-gRNA2 |
| DGP356 | pUC19-gRNA3 |
| DGP357 | pUC19-gRNA-negative |

# Chapter 5: Conclusion

## 5.1 Summary

In this dissertation, I have sought to understand the impact of fluctuating environments on the course and outcomes of adaptive evolution in *S. cerevisiae*. The selective pressure that fluctuating environments impose is complex and challenging to study. Here, I used chemostats to undertake experimental evolution which has the benefit of controlling for variables external to changes in nutrient content, such as growth rate and population size. By combining experimental evolution in the lab with molecular biology and next generation sequencing I was able to identify principles that govern selection in environments that fluctuate periodically in nutrient concentration and in nutrient quality.

In chapter 2, I addressed the question of how periodic fluctuations in nutrient concentrations impact populations with pre-existing genetic variation. I used a synthetic population comprising isogenic cells that differ by single gene deletions at all nonessential genes in *S. cerevisiae*. I fluctuated the concentration of both carbon and nitrogen reciprocally over four periods and tracked the frequencies of ~4000 unique genotypes in each population. I found that even over short-term selection, the effect of periodically fluctuating environments on population diversity is profound. In three periodically fluctuating environments, that different either in the magnitude of the fluctuation or in the quality of nutrient, increased diversity was maintained in comparison to two static environments. Previously, the effects of environmental fluctuations have been contradictory and ambiguous, however, our study offers a highly controlled

experiment that shows definitively that fluctuating selection results in the maintenance of greater genetic diversity in comparison to static conditions regardless of the magnitude of the fluctuation and over multiple types of nutrients.

I showed that the maintenance in diversity is driven by two main factors 1) weakened selection and 2) genotypes that oscillate in frequency of abundance. By performing mathematical modeling of the growth dynamics of each of the ~4000 genotypes I was able to identify several growth behaviors that are condition-dependent. Contrary to the classical approach in which microbial fitness is measured assuming monotonic behavior, when polynomial models were fit to growth, a large proportion of genotypes were found to follow either quadratic or cubic behavior. Surprisingly, I also found oscillating genotypes with a period that matched that of the fluctuation period uniquely in fluctuating environments. Fluctuating conditions harbored the greatest proportion of genotypes that showed neutral growth. These two behaviors resulted in the maintenance of a greater number of unique genotypes in the population which effectively maintained greater diversity. From this lab-controlled experiment with high-resolution sampling I was able to show how periodically fluctuating selection can result in the maintenance of genetic diversity and the forces that result in this phenomenon.

In Chapter 3, I performed a study to determine whether diversity is maintained under fluctuating selection over long-term adaptive evolution in which populations have an opportunity to generate *de novo* mutations. I focused on a single class of mutations,

copy number variants (CNVs), that are known to be under strong selection in static nutrient-limited chemostats. I expanded an established CNV-reporter assay by implementing the use of two fluorophores which enabled that tracking of copy number variation at two distinct genomic loci simultaneously. I constructed strains in which the amino acid permease GAP1 had mCitrine inserted adjacent to it and either the proline transporter PUT4 or the ammonium-sulfate transporter MEP2 had mCherry inserted adjacent to them. By performing long-term experimental evolution over 250 generations of these clonal populations in static and fluctuating conditions I was able to reveal dynamics of selection that were unique between conditions and genetic loci.

Fluctuating conditions resulted in less predictable and repeatable dynamics of CNVs whereas static conditions showed greater repeatability between replicate populations. I observed the generation and selection for genotypes that had amplifications at two different loci resulting in  CNV generalists in fluctuating conditions. Contrary to expectation, in static glutamine-limitation CNV generalists also arose to significant frequencies in the population, but this was not observed in the other two static conditions. These results suggest that fluctuating conditions select for greater diversity outcomes at least from the perspective of strategies that are CNV dependent, however, it remains unclear whether including the analysis of other mutations such as single nucleotide variants, would support this conclusion. Further experiments that include the use of whole population sequencing to identify other mutations in the population, as well as the use of synthetic DNA barcodes for lineage tracking, will reveal the effects of fluctuations on diversity during long-term experimental evolution.

In Chapter 4, I introduced a new DNA barcoding system for performing lineage tracking that is versatile, easy to use, and minimally invasive. I have established the experimental foundations for the system and mapped out the required future experiments to validate its success. I showed that the use of an inducible CRISPR/Cas9 system to increase the efficiency of incorporation of a complex DNA barcode library would be an improvement from currently existing technology both by decreasing the size of exogenous DNA sequence inserted by more than half, and by reducing the negative effects on fitness that are due to positional effects.

## 5.2 Future directions

In this thesis I have described the use of several important tools for the study of adaptive evolution in a laboratory setting. I have made improvements on these pre-existing tools which extend their utility for addressing questions that have previously been challenging to tackle. One of the largest challenges of the study of adaptive evolution in microbes is the ability to track genetic and phenotypic variation at the individual level. Next generation sequencing enables detection of genetic mutations that exist in a population, but without performing single-cell sequencing it is impossible to know which mutations exist within the same individual cell or come from the same lineage. Single-cell sequencing remains expensive and currently is limited by the number of single cells that can be simultaneously sequenced which makes it prohibitive to use to study adaptive changes which requires numerous samplings over time of populations with millions of individual cells. However, by combining multiple

technologies, future studies can extract single-cell information at a lower cost.

Using a dual-fluorescence CNV reporter during experimental evolution enables the accurate reporting on the gene copy number of two loci of interest simultaneously at low cost. As such combining this assay with multiplexed chemostat arrays enables the high-throughput testing of CNV evolution under a variety of conditions. It also allows for the application of sort-seq experiments in which fluorescence can be used as a marker for subpopulations with different gene copy numbers. Combined with FACS subpopulations can be isolated and further analyzed by NGS experiments such as whole genome sequencing or RNAseq to decipher which mutations are unique to this subpopulation or whether there are unique patterns of expression characterizing it.

Using a two-color system expands the questions that can be addressed about experimental evolution in microbes. While the focus in this thesis has been on understanding changes in gene copy number in haploid cells, this system has important applications in diploids as well. A dual color system can be used to detect changes in gene allele number in heterozygotes to understand how hybrids evolve and to observe the process of loss of heterozygosity for example and the implications of the environment in this process. Another application is aimed at understanding whether CNVs can have a direct and proportional effect on transcription or whether cells buffer DNA copy number changes. By using one fluorophore as a CNV-reporter at a locus of interest and a second fluorophore as a transcriptional reporter with the promoter of the locus of interest then it is possible to couple information on DNA copy number with

expression information to detect changes during experimental evolution.

While the focus of this thesis has been on microbial evolution the utility of this assay can easily be extended to other organisms and systems as well. CNVs are implicated in different types of cancers and the use of a dual-CNV reporter to understand tumor evolution would have important implications on developing regimes for the administration of therapeutics. Because of the versatility of fluorescent reporters and their use in many model organisms, the extension of this assay is expected to be seamless.

# REFERENCES


Airoldi, Edoardo M., Darach Miller, Rodoniki Athanasiadou, Nathan Brandt, Farah Abdul-Rahman, Benjamin Neymotin, Tatsu Hashimoto, Tayebeh Bahmani, and David Gresham. 2016. “Steady-State and Dynamic Gene Expression Programs in Saccharomyces Cerevisiae in Response to Variation in Environmental Nitrogen.” *Molecular Biology of the Cell* 27 (8): 1383–96.

Bell, Graham. 2008. *Selection*. The Mechanism of Evolution. Oxford Univ Press. 2010. “Fluctuating Selection: The Perpetual Renewal of Adaptation in Variable Environments.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 365 (1537): 87–97.

Bergland, Alan O., Emily L. Behrman, Katherine R. O’Brien, Paul S. Schmidt, and Dmitri A. Petrov. 2014. “Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila.” *PLoS Genetics* 10 (11): e1004775.

Berry, David B., and Audrey P. Gasch. 2008. “Stress-Activated Genomic Expression Changes Serve a Preparative Role for Impending Stress in Yeast.” *Molecular Biology of the Cell* 19 (11): 4580–87.

Bertram, Jason, and Joanna Masel. 2019. “Different Mechanisms Drive the Maintenance of Polymorphism at Loci Subject to Strong versus Weak Fluctuating Selection.” *Evolution; International Journal of Organic Evolution* 73 (5): 883–96.

Bigger, Joseph Warwick. 1944. *Treatment of Staphylococcal Infections with Penicillin by Intermittent Sterilisation*.

Bisson, L. F., L. Neigeborn, M. Carlson, and D. G. Fraenkel. 1987. “The SNF3 Gene Is Required for High-Affinity Glucose Transport in Saccharomyces Cerevisiae.”

*Journal of Bacteriology* 169 (4): 1656–62.

Blake, William J., Gábor Balázsi, Michael A. Kohanski, Farren J. Isaacs, Kevin F. Murphy, Yina Kuang, Charles R. Cantor, David R. Walt, and James J. Collins. 2006. “Phenotypic Consequences of Promoter-Mediated Transcriptional Noise.” *Molecular Cell* 24 (6): 853–65.

Blundell, Jamie R., Katja Schwartz, Danielle Francois, Daniel S. Fisher, Gavin J. Sherlock, and Sasha F. Levy. 2017. “The Dynamics of Adaptive Genetic Diversity during the Early Stages of Clonal Evolution.” *bioRxiv*. https://doi.org/10.1101/170589.

Boer, Viktor M., Christopher A. Crutchfield, Patrick H. Bradley, David Botstein, and Joshua D. Rabinowitz. 2010. “Growth-Limiting Intracellular Metabolites in Yeast Growing under Diverse Nutrient Limitations.” *Molecular Biology of the Cell* 21 (1): 198–211.

Boyce, Kylie J., Chengjun Cao, Chaoyang Xue, and Alexander Idnurm. 2020. “A Spontaneous Mutation in DNA Polymerase POL3 during in Vitro Passaging Causes a Hypermutator Phenotype in Cryptococcus Species.” *DNA Repair* 86 (February): 102751.

Brauer, Matthew J., Curtis Huttenhower, Edoardo M. Airoldi, Rachel Rosenstein, John C. Matese, David Gresham, Viktor M. Boer, Olga G. Troyanskaya, and David Botstein. 2008. “Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast.” *Molecular Biology of the Cell* 19 (1): 352–67.

Bresson, Stefan, Vadim Shchepachev, Christos Spanos, Tomasz W. Turowski, Juri Rappsilber, and David Tollervey. 2020. “Stress-Induced Translation Inhibition

through Rapid Displacement of Scanning Initiation Factors.” *Molecular Cell* 80 (3): 470–84.e8.

Buziol, Stefan, Lydia Warth, Ivana Magario, Andreas Freund, Martin Siemann-Herzberg, and Matthias Reuss. 2008. “Dynamic Response of the Expression of hxt1, hxt5 and hxt7 Transport Proteins in Saccharomyces Cerevisiae to Perturbations in the Extracellular Glucose Concentration.” *Journal of Biotechnology* 134 (3-4): 203–10.

Chantranupong, Lynne, Rachel L. Wolfson, and David M. Sabatini. 2015. “Nutrient-Sensing Mechanisms across Evolution.” *Cell* 161 (1): 67–83.

Chesson, P. 1994. “Multispecies Competition in Variable Environments.” *Theoretical Population Biology*. https://doi.org/10.1006/tpbi.1994.1013.

Childs, Dylan Z., C. J. E. Metcalf, and Mark Rees. 2010. “Evolutionary Bet-Hedging in the Real World: Empirical Evidence and Challenges Revealed by Plants.” *Proceedings. Biological Sciences / The Royal Society* 277 (1697): 3055–64.

Connallon, Tim, and Andrew G. Clark. 2015. “The Distribution of Fitness Effects in an Uncertain World.” *Evolution*. https://doi.org/10.1111/evo.12673.

Cooper, Tim F., and Richard E. Lenski. 2010. “Experimental Evolution with E. Coli in Diverse Resource Environments. I. Fluctuating Environments Promote Divergence of Replicate Populations.” *BMC Evolutionary Biology* 10 (1): 11.

Cvijović, Ivana, Benjamin H. Good, Elizabeth R. Jerison, and Michael M. Desai. 2015. “Fate of a Mutation in a Fluctuating Environment.” *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1505406112.

Dahiya, Rashmi, Qing Hu, and Peter Ly. 2021. “Mechanistic Origins of Diverse Genome Rearrangements in Cancer.” *Seminars in Cell & Developmental Biology*, April.

https://doi.org/10.1016/j.semcdb.2021.03.003.

Day, Troy, Todd Parsons, Amaury Lambert, and Sylvain Gandon. 2020. “The Price Equation and Evolutionary Epidemiology.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 375 (1797): 20190357.

Dean, Antony M. 2005. “Protecting Haploid Polymorphisms in Temporally Variable Environments.” *Genetics* 169 (2): 1147–56.

Delneri, Daniela. 2010. “Barcode Technology in Yeast: Application to Pharmacogenomics.” *FEMS Yeast Research* 10 (8): 1083–89.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. “Rates of Spontaneous Mutation.” *Genetics* 148 (4): 1667–86.

Eelderink-Chen, Zheng, Gabriella Mazzotta, Marcel Sturre, Jasper Bosman, Till Roenneberg, and Martha Merrow. 2010. “A Circadian Clock in Saccharomyces Cerevisiae.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (5): 2043–47.

Eyre-Walker, Adam, and Peter D. Keightley. 2007. “The Distribution of Fitness Effects of New Mutations.” *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2146.

Frank, Steven A. 2013. “Evolution of Robustness and Cellular Stochasticity of Gene Expression.” *PLoS Biology* 11 (6): e1001578.

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. 2000. “Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes.” *Molecular Biology of the Cell* 11 (12): 4241–57.

Gietz, R. Daniel, and Robert H. Schiestl. 2007. “Large-Scale High-Efficiency Yeast

Transformation Using the LiAc/SS Carrier DNA/PEG Method." *Nature Protocols* 2 (1): 38–41.

Good, Benjamin H., Michael J. McDonald, Jeffrey E. Barrick, Richard E. Lenski, and Michael M. Desai. 2017. "The Dynamics of Molecular Evolution over 60,000 Generations." *Nature* 551 (7678): 45–50.

Gresham, David, Michael M. Desai, Cheryl M. Tucker, Harry T. Jenq, Dave A. Pai, Alexandra Ward, Christopher G. DeSevo, David Botstein, and Maitreya J. Dunham. 2008. "The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast." *PLoS Genetics* 4 (12): e1000303.

Harari, Yaniv, Yoav Ram, and Martin Kupiec. 2018. "Frequent Ploidy Changes in Growing Yeast Cultures." *Current Genetics* 64 (5): 1001–4.

Healey, David, Kevin Axelrod, and Jeff Gore. 2016. "Negative Frequency-Dependent Interactions Can Underlie Phenotypic Heterogeneity in a Clonal Microbial Population." *Molecular Systems Biology* 12 (8): 877.

Hedrick, Philip W. 2007. "Balancing Selection." *Current Biology: CB* 17 (7): R230–31.

Hietpas, Ryan T., Claudia Bank, Jeffrey D. Jensen, and Daniel N. A. Bolon. 2013. "Shifting Fitness Landscapes in Response to Altered Environments." *Evolution; International Journal of Organic Evolution* 67 (12): 3512–22.

Hoffman, C. S., and F. Winston. 1987. "A Ten-Minute DNA Preparation from Yeast Efficiently Releases Autonomous Plasmids for Transformation of Escherichia Coli." *Gene* 57 (2-3): 267–72.

Hong, Jungeui, Nathan Brandt, Farah Abdul-Rahman, Ally Yang, Tim Hughes, and David Gresham. 2018. "An Incoherent Feedforward Loop Facilitates Adaptive

Tuning of Gene Expression." *eLife* 7 (April). https://doi.org/10.7554/eLife.32323.

Hong, Jungeui, and David Gresham. 2014. "Molecular Specificity, Convergence and Constraint Shape Adaptive Evolution in Nutrient-Poor Environments." *PLoS Genetics* 10 (1): e1004041.

Johnson, Carl Hirschie, and Susan S. Golden. 2006. "Circadian Rhythms in Cyanobacteria." *eLS*. https://doi.org/10.1038/npg.els.0004291.

Kang, Minjeong, Kangsan Kim, Donghui Choe, Suhyung Cho, Sun Chang Kim, Bernhard Palsson, and Byung-Kwan Cho. 2019. "Inactivation of a Mismatch-Repair System Diversifies Genotypic Landscape of During Adaptive Laboratory Evolution." *Frontiers in Microbiology* 10 (August): 1845.

Kashtan, Nadav, Elad Noor, and Uri Alon. 2007. "Varying Environments Can Speed up Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 104 (34): 13711–16.

Kerr, Benjamin, Margaret A. Riley, Marcus W. Feldman, and Brendan J. M. Bohannan. 2002. "Local Dispersal Promotes Biodiversity in a Real-Life Game of Rock-Paper-Scissors." *Nature* 418 (6894): 171–74.

Kubitschek, Herbert Ernest. 1970. *Introduction to Research with Continuous Cultures*. Prentice Hall.

Kvitek, Daniel J., and Gavin Sherlock. 2011. "Reciprocal Sign Epistasis between Frequently Experimentally Evolved Adaptive Mutations Causes a Rugged Fitness Landscape." *PLoS Genetics* 7 (4): e1002056.

Lambert, Guillaume, and Edo Kussell. 2014. "Memory and Fitness Optimization of Bacteria under Fluctuating Environments." *PLoS Genetics* 10 (9): e1004556.

Landis, Elizabeth A., Angela M. Oliverio, Erin A. McKenney, Lauren M. Nichols, Nicole Kfoury, Megan Biango-Daniels, Leonora K. Shell, et al. 2021. “The Diversity and Function of Sourdough Starter Microbiomes.” *eLife* 10 (January). https://doi.org/10.7554/eLife.61644.

Lauer, Stephanie, Grace Avecilla, Pieter Spealman, Gunjan Sethia, Nathan Brandt, Sasha F. Levy, and David Gresham. 2018. “Single-Cell Copy Number Variant Detection Reveals the Dynamics and Diversity of Adaptation.” *PLoS Biology* 16 (12): e3000069.

Letten, A. D., M. K. Dhami, P. J. Ke, and T. Fukami. 2018. “Species Coexistence through Simultaneous Fluctuation-Dependent Mechanisms.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (26). https://doi.org/10.1073/pnas.1801846115.

Levins, Richard. 1968. *Evolution in Changing Environments: Some Theoretical Explorations*.

Levy, Sasha F., Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. 2015. “Quantitative Evolutionary Dynamics Using High-Resolution Lineage Tracking.” *Nature* 519 (7542): 181–86.

Levy, Sasha F., Naomi Ziv, and Mark L. Siegal. 2012. “Bet Hedging in Yeast by Heterogeneous, Age-Correlated Expression of a Stress Protectant.” *PLoS Biology* 10 (5): e1001325.

Lin, Wei-Hsiang, and Edo Kussell. 2016. “Complex Interplay of Physiology and Selection in the Emergence of Antibiotic Resistance.” *Current Biology: CB* 26 (11): 1486–93.

Li, Yuping, Sandeep Venkataram, Atish Agarwala, Barbara Dunn, Dmitri A. Petrov, Gavin Sherlock, and Daniel S. Fisher. 2018. “Hidden Complexity of Yeast Adaptation under Simple Evolutionary Conditions.” *Current Biology: CB* 28 (4): 515–25.e6.

Lorenz, M. C., and J. Heitman. 1998. “The MEP2 Ammonium Permease Regulates Pseudohyphal Differentiation in Saccharomyces Cerevisiae.” *The EMBO Journal* 17 (5): 1236–47.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550.

Lyons, Eli, Paul Sheridan, Georg Tremmel, Satoru Miyano, and Sumio Sugano. 2017. “Large-Scale DNA Barcode Library Generation for Biomolecule Identification in High-Throughput Screens.” *Scientific Reports* 7 (1): 13899.

Magditch, Denise A., Tong-Bao Liu, Chaoyang Xue, and Alexander Idnurm. 2012. “DNA Mutations Mediate Microevolution between Host-Adapted Forms of the Pathogenic Fungus Cryptococcus Neoformans.” *PLoS Pathogens* 8 (10): e1002936.

Mansisidor, Andrés, Temistocles Molinar Jr, Priyanka Srivastava, Demetri D. Dartis, Adriana Pino Delgado, Hannah G. Blitzblau, Hannah Klein, and Andreas Hochwagen. 2018. “Genomic Copy-Number Loss Is Rescued by Self-Limiting Production of DNA Circles.” *Molecular Cell* 72 (3): 583–93.e4.

McKinney, Jennifer A., Guliang Wang, Anirban Mukherjee, Laura Christensen, Sai H. Sankara Subramanian, Junhua Zhao, and Karen M. Vasquez. 2020. “Distinct DNA Repair Pathways Cause Genomic Instability at Alternative DNA Structures.” *Nature*

*Communications* 11 (1): 236.

Miller, Aaron W., Corrie Befort, Emily O. Kerr, and Maitreya J. Dunham. 2013. “Design and Use of Multiplexed Chemostat Arrays.” *Journal of Visualized Experiments: JoVE*, no. 72 (February): e50262.

Miller, Darach, Nathan Brandt, and David Gresham. 2018. “Systematic Identification of Factors Mediating Accelerated mRNA Degradation in Response to Changes in Environmental Nitrogen.” *PLoS Genetics* 14 (5): e1007406.

Monod, Jacques. 1950. “La Technique de Culture Continue: Théorie et Applications.” *Annales de l’Institut Pasteur, Paris*, Selected Papers in Molecular Biology by Jacques Monod, , no. 79 (January): 390–410.

Morand, Serge, and Claire Lajaunie. 2018. “Loss of Biological Diversity and Emergence of Infectious Diseases.” *Biodiversity and Health*. https://doi.org/10.1016/b978-1-78548-115-4.50003-2.

Morley, Valerie J., and Paul E. Turner. 2017. “Dynamics of Molecular Evolution in RNA Virus Populations Depend on Sudden versus Gradual Environmental Change.” *Evolution; International Journal of Organic Evolution* 71 (4): 872–83.

Mukherji, Atish, Ahmad Kobiita, Tao Ye, and Pierre Chambon. 2013. “Homeostasis in Intestinal Epithelium Is Orchestrated by the Circadian Clock and Microbiota Cues Transduced by TLRs.” *Cell* 153 (4): 812–27.

Nelson, Paul, and Joanna Masel. 2018. “Evolutionary Capacitance Emerges Spontaneously during Adaptation to Environmental Changes.” *Cell Reports* 25 (1): 249–58.

New, Aaron M., Bram Cerulus, Sander K. Govers, Gemma Perez-Samper, Bo Zhu,

Sarah Boogmans, Joao B. Xavier, and Kevin J. Verstrepen. 2014. “Different Levels of Catabolite Repression Optimize Growth in Stable and Variable Environments.” *PLoS Biology* 12 (1): e1001764.

Nguyen, Jen, Juanita Lara-Gutiérrez, and Roman Stocker. 2020. “Environmental Fluctuations and Their Effects on Microbial Communities, Populations, and Individuals.” *FEMS Microbiology Reviews*, December. https://doi.org/10.1093/femsre/fuaa068.

Novick, A., and L. Szilard. 1950. “Description of the Chemostat.” *Science* 112 (2920): 715–16.

Olofsson, Helen, Jörgen Ripa, and Niclas Jonzén. 2009. “Bet-Hedging as an Evolutionary Game: The Trade-off between Egg Size and Number.” *Proceedings. Biological Sciences / The Royal Society* 276 (1669): 2963–69.

Ozcan, S., J. Dover, A. G. Rosenwald, S. Wölfl, and M. Johnston. 1996. “Two Glucose Transporters in Saccharomyces Cerevisiae Are Glucose Sensors That Generate a Signal for Induction of Gene Expression.” *Proceedings of the National Academy of Sciences of the United States of America* 93 (22): 12428–32.

Payen, Celia, Sara C. Di Rienzi, Giang T. Ong, Jamie L. Pogachar, Joseph C. Sanchez, Anna B. Sunshine, M. K. Raghuraman, Bonita J. Brewer, and Maitreya J. Dunham. 2014. “The Dynamics of Diverse Segmental Amplifications in Populations of Saccharomyces Cerevisiae Adapting to Strong Selection.” *G3* 4 (3): 399–409.

Persson, Sebastian, Niek Welkenhuysen, Sviatlana Shashkova, and Marija Cvijovic. 2020. “Fine-Tuning of Energy Levels Regulates via a SNF1-Dependent Feedback Loop.” *Frontiers in Physiology* 11 (August): 954.

Razinkov, Ivan A., Bridget L. Baumgartner, Matthew R. Bennett, Lev S. Tsimring, and Jeff Hasty. 2013. “Measuring Competitive Fitness in Dynamic Environments.” *The Journal of Physical Chemistry. B* 117 (42): 13175–81.

Rintala, Eija, Marilyn G. Wiebe, Anu Tamminen, Laura Ruohonen, and Merja Penttilä. 2008. “Transcription of Hexose Transporters of Saccharomyces Cerevisiae Is Affected by Change in Oxygen Provision.” *BMC Microbiology* 8 (March): 53.

Robertson, J. Brian, Chris C. Stowers, Erik Boczko, and Carl Hirschie Johnson. 2008. “Real-Time Luminescence Monitoring of Cell-Cycle and Respiratory Oscillations in Yeast.” *Proceedings of the National Academy of Sciences of the United States of America* 105 (46): 17988–93.

Robinson, David G., Wei Chen, John D. Storey, and David Gresham. 2013. “Design and Analysis of Bar-Seq Experiments.” *G3* 4 (1): 11–18.

Ronen, Michal, and David Botstein. 2006. “Transcriptional Response of Steady-State Yeast Cultures to Transient Perturbations in Carbon Source.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (2): 389–94.

Rosche, W. A., and P. L. Foster. 1999. “The Role of Transient Hypermutators in Adaptive Mutation in Escherichia Coli.” *Proceedings of the National Academy of Sciences of the United States of America* 96 (12): 6862–67.

Saldanha, Alok J., Matthew J. Brauer, and David Botstein. 2004. “Nutritional Homeostasis in Batch and Steady-State Culture of Yeast.” *Molecular Biology of the Cell* 15 (9): 4089–4104.

Salignon, Jérôme, Magali Richard, Etienne Fulcrand, Hélène Duplus-Bottin, and Gaël Yvert. 2018. “Genomics of Cellular Proliferation in Periodic Environmental

Fluctuations.” *Molecular Systems Biology* 14 (3): e7823.

Salim, Devika, William D. Bradford, Boris Rubinstein, and Jennifer L. Gerton. 2021. “DNA Replication, Transcription, and H3K56 Acetylation Regulate Copy Number and Stability at Tandem Repeats.” *G3* , March. https://doi.org/10.1093/g3journal/jkab082.

Schlecht, Ulrich, Zhimin Liu, Jamie R. Blundell, Robert P. St Onge, and Sasha F. Levy. 2017. “A Scalable Double-Barcode Sequencing Platform for Characterization of Dynamic Protein-Protein Interactions.” *Nature Communications* 8 (May): 15586.

Schlomann, Brandon H., and Raghuveer Parthasarathy. 2019. “Timescales of Gut Microbiome Dynamics.” *Current Opinion in Microbiology* 50 (August): 56–63.

Smukowski Heil, Caiti S., Christopher R. L. Large, Kira Patterson, Angela Shang-Mei Hickey, Chiann-Ling C. Yeh, and Maitreya J. Dunham. 2019. “Temperature Preference Can Bias Parental Genome Retention during Hybrid Evolution.” *PLoS Genetics* 15 (9): e1008383.

Sniegowski, P. D., P. J. Gerrish, and R. E. Lenski. 1997. “Evolution of High Mutation Rates in Experimental Populations of E. Coli.” *Nature* 387 (6634): 703–5.

Spealman, Pieter, Jaden Burrell, and David Gresham. 2020. “Inverted Duplicate DNA Sequences Increase Translocation Rates through Sequencing Nanopores Resulting in Reduced Base Calling Accuracy.” *Nucleic Acids Research* 48 (9): 4940–45.

Spies, Daniel, Peter F. Renz, Tobias A. Beyer, and Constance Ciaudo. 2019. “Comparative Analysis of Differential Gene Expression Tools for RNA Sequencing Time Course Data.” *Briefings in Bioinformatics*. https://doi.org/10.1093/bib/bbx115.

Subramanian, Aravind, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, et al. 2005. “Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (43): 15545–50.

Sun, Siyu, Anastasia Baryshnikova, Nathan Brandt, and David Gresham. 2020. “Genetic Interaction Profiles of Regulatory Kinases Differ between Environmental Conditions and Cellular States.” *Molecular Systems Biology* 16 (5): e9167.

Tagkopoulos, Ilias, Yir-Chung Liu, and Saeed Tavazoie. 2008. “Predictive Behavior within Microbial Genetic Networks.” *Science* 320 (5881): 1313–17.

Tan, Jiaqi, Jennifer B. Rattray, Xian Yang, and Lin Jiang. 2017. “Spatial Storage Effect Promotes Biodiversity during Adaptive Radiation.” *Proceedings. Biological Sciences / The Royal Society* 284 (1858). https://doi.org/10.1098/rspb.2017.0841.

Thaiss, Christoph A., Maayan Levy, Tal Korem, Lenka Dohnalová, Hagit Shapiro, Diego A. Jaitin, Eyal David, et al. 2016. “Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations.” *Cell* 167 (6): 1495–1510.e12.

Thaiss, Christoph A., David Zeevi, Maayan Levy, Gili Zilberman-Schapira, Jotham Suez, Anouk C. Tengeler, Lior Abramson, et al. 2014. “Transkingdom Control of Microbiota Diurnal Oscillations Promotes Metabolic Homeostasis.” *Cell* 159 (3): 514–29.

Thompson, Dawn A., Michael M. Desai, and Andrew W. Murray. 2006. “Ploidy Controls the Success of Mutators and Nature of Mutations during Budding Yeast Evolution.” *Current Biology: CB* 16 (16): 1581–90.

Tu, Benjamin P., Andrzej Kudlicki, Maga Rowicka, and Steven L. McKnight. 2005. “Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes.” *Science* 310 (5751): 1152–58.

Uschner, Friedemann, and Edda Klipp. 2014. “Information Processing in the Adaptation of Saccharomyces Cerevisiae to Osmotic Stress: An Analysis of the Phosphorelay System.” *Systems and Synthetic Biology* 8 (4): 297–306.

VanderSluis, Benjamin, David C. Hess, Colin Pesyna, Elias W. Krumholz, Tahin Syed, Balázs Szappanos, Corey Nislow, et al. 2014. “Broad Metabolic Sensitivity Profiling of a Prototrophic Yeast Deletion Collection.” *Genome Biology* 15 (4): R64.

Veening, Jan-Willem, Wiep Klaas Smits, and Oscar P. Kuipers. 2008. “Bistability, Epigenetics, and Bet-Hedging in Bacteria.” *Annual Review of Microbiology* 62: 193–210.

Velasco, Silvia, Mahmoud M. Ibrahim, Akshay Kakumanu, Görkem Garipler, Begüm Aydin, Mohamed Ahmed Al-Sayegh, Antje Hirsekorn, et al. 2017. “A Multi-Step Transcriptional and Chromatin State Cascade Underlies Motor Neuron Programming from Embryonic Stem Cells.” *Cell Stem Cell* 20 (2): 205–17.e8.

Venkataram, Sandeep, Barbara Dunn, Yuping Li, Atish Agarwala, Jessica Chang, Emily R. Ebel, Kerry Geiler-Samerotte, et al. 2016. “Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast.” *Cell* 166 (6): 1585–96.e22.

Walworth, Nathan G., Emily J. Zakem, John P. Dunne, Sinéad Collins, and Naomi M. Levine. 2020. “Microbial Evolutionary Strategies in a Dynamic Ocean.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (11):

5943–48.

Wang, Haifeng, Marie La Russa, and Lei S. Qi. 2016. “CRISPR/Cas9 in Genome Editing and Beyond.” *Annual Review of Biochemistry* 85 (June): 227–64.

Wang, Jue, Esha Atolia, Bo Hua, Yonatan Savir, Renan Escalante-Chong, and Michael Springer. 2015. “Natural Variation in Preparation for Nutrient Depletion Reveals a Cost-Benefit Tradeoff.” *PLoS Biology* 13 (1): e1002041.

West, Robert, and Mauro Mobilia. 2020. “Fixation Properties of Rock-Paper-Scissors Games in Fluctuating Populations.” *Journal of Theoretical Biology* 491 (April): 110135.

Wiser, Michael J., and Richard E. Lenski. 2015. “A Comparison of Methods to Measure Fitness in Escherichia Coli.” *PloS One* 10 (5): e0126210.

# Fluctuating environments maintain genetic diversity through neutral fitness effects and balancing selection - Supplemental Methods

Abdul-Rahman F, Tranchina D and Gresham D

February 2021

## S1 Mathematical and Statistical Methods

### S1.1 Symbols and notation

Symbols, notation and mathematical expressions used in this section are summarized in Table S1.

### S1.2 Basis of barcode count normalization

#### S1.2.1 Dynamics of genotypic populations

In the chemostat, a genetically homogeneous population of size $n$ grows through cell division at a rate given by $n'(t) = \lambda(t)\, n(t)$, where $\lambda(t)$ is the instantaneous, *time-dependent*, growth rate per cell; i.e., $\lambda(t) = n'(t)/n(t)$. We assume here that there is no cell death. If there were, $\lambda$ could be taken to mean the net rate of change per cell due to cell division and cell death. Each genotypic population in the chemostat is identified by its unique DNA barcode. Our normalization method allows us to estimate, for each genotypic population, the time average of its growth rate per cell (fitness), $\lambda_i(t)$, between any 2 time points, minus the arithmetic mean over all genotypes. Furthermore, with generalized linear model fitting, we will be able to estimate $\lambda_i(t)$ itself, minus the arithmetic mean. The normalization method is based on the population dynamics below.

In the chemostat, with $m$ genotypes, the number of cells with genotype $i$ at time $t$, $n_i(t)$, changes according to

$$\frac{d}{dt} n_i = [\lambda_i(t) - \beta]\, n_i, \text{ or} \tag{S1a}$$

$$\frac{d}{dt} \log(n_i) = \lambda_i(t) - \beta, \tag{S1b}$$

| Expression | Definition |
|---|---|
| $\lambda_i(t)$ | instantaneous growth rate per cell and our measure of fitness at time $t$ for sub-population with genotype $i$, as identified by its unique DNA barcode |
| $\Lambda_i(t) \,=\, \int_0^t \lambda_i(t')\,dt'$ | integrated fitness, for genotype $i$, over time interval $(0,t)$ |
| $(1/t)\int_0^t \lambda_i(t')\,dt'$ | temporal mean of fitness, for genotype $i$, over time interval $(0,t)$ |
| $m$ | number of genotypes in the chemostat |
| $\mu_\lambda^{\mathrm{a}}(t) \,=\, \frac{1}{m}\sum_{k=1}^m \lambda_k(t)$ | arithmetic mean, over genotypes, of instantaneous fitness at time $t$ |
| $\mu_\Lambda^{\mathrm{a}}(t) \,=\, \frac{1}{m}\sum_{k=1}^m \Lambda_k(t)$ | arithmetic mean of integrated fitness over time interval $(0,t)$ |
| $n_i(t)$ | size (or density) of sub-population with genotype $i$ at time $t$ |
| $N(t) \,=\, \sum_{i=1}^m n_i(t)$ | total number of cells (or density) in the chemostat at time $t$ |
| $p_i(t)$ | proportion of cells of genotype $i$ at time $t$ |
| $\lambda_i^{\mathrm{rel}}(t) \,=\, \lambda_i(t) \,-\, \mu_\lambda^{\mathrm{a}}(t)$ | relative (to arithmetic mean) fitness of genotype $i$ at time $t$ |
| $\Lambda_i^{\mathrm{rel}}(t) \,=\, \Lambda_i(t) \,-\, \mu_\Lambda^{\mathrm{a}}(t)$ | relative integrated fitness for genotype $i$ over time interval $(0,t)$ |
| $(1/t)\,\Lambda_i^{\mathrm{rel}}(t)$ | temporal mean of relative fitness for genotype $i$ over time interval $(0,t)$ |
| $\mu_\lambda(t) \,=\, \sum_{i=1}^m \lambda_i(t)\,p_i(t)$ | population mean of fitness at time $t$ |
| $\sigma_\lambda^2(t) \,=\, \sum_{i=1}^m p_i(t)\,\big[\lambda_i(t) \,-\, \mu_\lambda(t)\big]^2$ | population variance of fitness at time $t$ |

Table S1: Definition of symbols and mathematical expressions

where $\beta$ is the dilution rate constant.

In our study, the growth rate per cell $\lambda_i(t)$ is empirical. Our experimental plots of log normalized barcode count versus time (Figure 3) show a rich variety of time courses in which many growth rates per cell are clearly not constant. We show below in our computational modeling (S1.6.2) that an extension of standard mathematical model for the chemostat [10] with heterogeneous populations [2], in which growth rate per cell depends instantaneously on the limiting nutrient concentration, $S(t)$, cannot account for the complex dynamics of cell numbers that we observe here (see Figure 3 and S1.6.2).

In our normalization method, the dilution rate constant $\beta$ plays no role. Ignoring $\beta$ and integrating Eq. (S1a) gives

$$n_i(t) \,=\, n_i(0)\,\exp\big\{\Lambda_i(t)\big\},\ \text{where} \tag{S2a}$$

$$\Lambda_i(t) \,=\, \int_0^t \lambda_i(t')\,dt'. \tag{S2b}$$

Because $\Lambda_i(t)$ in Eq. (S2b) is the integral of the instantaneous fitness (growth rate per cell) for genotype $i$, $(1/t)\Lambda_i(t)$ is the temporal mean fitness over the time interval $(0,t)$.

The instantaneous fitness $\lambda_i(t)$ and the time average fitness $(1/t)\,\Lambda_i(t)$ are quantities we would like to estimate directly form our sequencing counts, but we don't know of any normalization method that can isolate these quantities. However, our normalization will allow us to determine for each genotype the fitness relative to the arithmetic mean over all genotypes. We define the relative fitness, $\lambda_i^{\text{rel}}(t)$, as

$$\lambda_i^{\text{rel}}(t) \stackrel{\text{def}}{=} \lambda_i(t) - \frac{1}{m}\sum_{k=1}^{m}\lambda_k(t), \tag{S3}$$

and the time-average relative fitness, $(1/t)\Lambda_i^{\text{rel}}(t)$, as

$$\frac{1}{t}\Lambda_i^{\text{rel}}(t) \stackrel{\text{def}}{=} \frac{1}{t}\left[\Lambda_i(t) - \frac{1}{m}\sum_{k=1}^{m}\Lambda_k(t)\right] \tag{S4}$$

### S1.2.2 Dynamics of barcode tags

We assume that an effluent-sample from the chemostat that is used for sequencing has the same proportions of populations as those found in the chemostat.

We combine up- and down-tags [3] in our count of tags from each genotype. The proportion of barcode tags corresponding to genotype $i$ depends on all $n_i(t)$ and all "yield coefficients", $\alpha_k$, for $k = 1, 2, \ldots, m$, where $m$ is the total number of genotypes. The yield coefficient $0 \leq \alpha_i \leq 2$ can be thought of the average number of tags per cell contributed by a cell in population $i$. Based on our counts of up- and down-tags for each genotype in each sequencing run, and sometimes finding discrepant values, it appears that the $\alpha_i$ differ from DNA barcode to DNA barcode. In our method of normalization and inference relative genotype abundance across time, the $\alpha_i$ will play no role. However, they do contribute conceptually to the development of our normalization method.

The average average number of detected barcode tags per cell depends on the probability per cell of 2 non-disjoint events, $U_i$ — the up-tag from a cell in population $i$ is detected, $D_i$ — the down-tag is detected — and their complements, $U_i'$ and $D_i'$. One tag is captured in the event $(U_i \cap D_i')$; similarly, one tag is captured in the event $(D_i \cap U_i')$; and two tags are detected in the event $(U_i \cap D_i)$. Consequently, $\alpha_i = 1 \cdot P(U_i \cap D_i') + 1 \cdot P(D_i \cap U_i') + 2 \cdot P(U_i \cap D_i) = P(U_i) + P(D_i)$.

We assume that the number of sequencing tags from population $i$ in a sequencing sample depends on the proportion of detectable tags $i$ among the whole population of detectable tags in the sequenced sample, and on the library size $\mathcal{L}$.

With these assumptions, the expected number of barcode tags from population $i$ in a library at time $t$, $\mu_i(t)$, is given by

$$\mu_i(t) = \frac{\alpha_i n_i(t)}{\sum_{k=1}^{m}\alpha_k n_k(t)}\,\mathcal{L}. \tag{S5}$$

The library size, $\mathcal{L}$ whose estimate is not relevant in our normalization method. In analogy with RNA-seq methods [4, 5, 9], we model the random integer number of detected tags from population $i$, $Y_i$, as a negative

binomial random variable with mean $\mu_i$ and a size parameter $a_i$, which is estimated in DESeq2 [4]. The motivation is that sampling noise alone would make the joint probability mass function of $[Y_1, Y_2, \ldots, Y_m]$ multinomial, which is very well approximated by a product of independent Poisson probability mass functions, as long as no particular genotype accounts for a sizable fraction of tags. However, in addition to the sampling noise, there are other sources of noise, such as the random number tags from populations of low number, stemming from the random number of such cells, deviations of $\alpha_i$ from the population average from replicate to replicate for fixed $i$, and experimental noise. In this case, the additional sources of noise imply that an overdispersed Poisson model is in order. The obvious choice, in analogy with RNA-seq, is the negative binomial.

## S1.3 Normalization of barcode counts

We propose a normalization of counts, $y_i$, for genotype $i$ in a library, by a factor, $s(t)$, that is given by the geometric mean of counts over all genotypes in the library. This normalization method does not give a quantity for each genotype that is directly related to its proportion in the overall population. Such a normalization method is not appropriate for inference of instantaneous growth rates per cell (fitness), or time-average fitness. The reason is that, while the numerator term in a proportion is determined in simple way as a function of time that is dependent on the instantaneous fitness, the denominator term is a function of time that depends on each of the individual population sizes as a function of time, rather than constant.

As shown below, the utility of our normalization is that it allows one to easily compute, for each genotype $i$, the time average fitness between any 2 time points, minus the arithmetic mean of this quantity over all genotypes in Eq. (S4). We refer to this as the relative temporal-mean fitness, where the term *relative* signifies relative to the arithmetic mean over genotypes. The same sort of definition of terms applies to instantaneous fitness and instantaneous relative fitness.

The reasoning behind our normalization stems from the fact that the geometric mean, over all genotypes, of population sizes, changes as an exponential function of time. To illustrate this idea, we consider the special case of growth rate constants per cell that are independent of time, and we ignore dilution. The well

known argument is as follows

$$\left[\prod_{j=1}^{m}\frac{n_j(t)}{n_j(0)}\right]^{1/m} = \left[\prod_{k=1}^{m}\exp\{\lambda_k\,t\}\right]^{1/m} \tag{S6a}$$

$$= \left[\exp\left(\sum_{k=1}^{m}\lambda_k\,t\right)\right]^{1/m} \tag{S6b}$$

$$= \exp\left(\left[\frac{1}{m}\sum_{k=1}^{m}\lambda_k\right]t\right) \tag{S6c}$$

$$= \exp\left(\mu_\lambda^{\mathrm{a}}\,t\right), \tag{S6d}$$

where $\mu_\lambda^{\mathrm{a}}$ is defined implicitly by Eqs. (S6c) and (S6d) as the arithmetic mean of $\lambda_k$ over all $k$ (genotypes).

In the case where we have time-dependent growth rates per cell, $\lambda_i(t)$, and dilution rate, $\beta$, the geometric mean of the population sizes is given by

$$\left[\prod_{k=1}^{m}\frac{n_k(t)}{n_k(0)}\right]^{1/m} = \exp\left(\frac{1}{m}\sum_{k=1}^{m}\Lambda_k(t)\;-\;\beta\,t\right) \tag{S7a}$$

$$= \exp\left(\mu_\Lambda^{\mathrm{a}}(t)\;-\;\beta\,t\right), \tag{S7b}$$

where $\mu_\Lambda^{\mathrm{a}}(t)$ is defined implicitly by Eqs. (S7a) and (S7b) at the arithmetic mean of $\Lambda_k(t)$ over all $k$. Recall that $\Lambda_k(t)$ is the integrated fitness of population $k$ over the time interval $(0,t)$.

What do we get when we normalize counts $y_i(t)$ by

$$s(t) \stackrel{\text{def}}{=} \left[\prod_{k=1}^{m}y_k(t)\right]^{1/m}, \tag{S8}$$

the geometric mean over $k$ of counts $y_k(t)$? To answer this question, first, we treat the normalization factor, $s(t)$, as a known parameter, rather than a random variable. This is the customary treatment of normalization factors in high throughput sequencing (e.g., DESeq2 [4], edgeR [5, 9], and limma [8]). Second, for sake of exposition only, we go further and replace $s(t)$ by the geometric mean over $k$ of the expected values of the counts, $\mu_k(t)$, rather than the actual counts. With this normalization, Eq. (S2a) for $n_i(t)$, and Eq. (S5) for

$\mu_i(t)$, the normalized count simplifies, and it has an expected value given by

$$\frac{\mu_{_i}(t)}{s(t)} \approx \frac{\mu_i(t)}{\left[\prod_{k=1}^{m} \mu_k(t)\right]^{1/m}} \tag{S9a}$$

$$= \frac{\alpha_i\, n_i(t)}{\left[\prod_{k=1}^{m} \alpha_k\, n_k(t)\right]^{1/m}} \tag{S9b}$$

$$= \frac{\alpha_i\, n_i(0)}{\left[\prod_{k=1}^{m} \alpha_k\, n_k(0)\right]^{1/m}} \exp\left[\Lambda_i(t) \,-\, \mu_\Lambda^{\mathrm{a}}(t)\right] \tag{S9c}$$

$$= \exp\left(\gamma_i + \Lambda_i(t) \,-\, \mu_\Lambda^{\mathrm{a}}(t)\right), \tag{S9d}$$

where

$$\gamma_i \stackrel{\mathrm{def}}{=} \log \frac{\alpha_i\, n_i(0)}{\left[\prod_{k=1}^{m} \alpha_k\, n_k(0)\right]^{1/m}} \tag{S10}$$

is the log of the normalized count at $t = 0$.

#### S1.3.1 Time-average relative fitness and inter-population differences between time-average fitness

It is convenient to give the expected normalized count for population $i$ a name,

$$z_i(t) \stackrel{\mathrm{def}}{=} \frac{\mu_{_i}(t)}{s(t)}. \tag{S11}$$

From Eq. (S9d) one can see that the time-average relative fitness over time interval $(0, t)$, as defined in Eq. (S4), is given by

$$\begin{aligned} \frac{1}{t}\Lambda_i^{\mathrm{rel}}(t) &\stackrel{\mathrm{def}}{=} \frac{1}{t}\left[\Lambda_i(t) \,-\, \mu_\Lambda^{\mathrm{a}}(t)\right] \\ &= \frac{\log z_i(t) \,-\, \log z_i(0)}{t}. \end{aligned} \tag{S12}$$

Similarly, the time-average relative fitness over any time interval $t_1 < t < t_2$ is given by

$$\frac{\log z_i(t_2) \,-\, \log z_i(t_1)}{t_2 \,-\, t_1},$$

the difference of log normalized count at times $t_2$ and $t_1$, divided by the time difference. The instantaneous relative fitness for genotype $i$ is given by

$$\lambda_i^{\mathrm{rel}}(t) \,=\, \frac{d}{dt} \log z_i(t). \tag{S13}$$

One can compute the difference between the time-average fitness for 2 genotypes, $i$ and $j$, from their relative time-average fitness (because the arithmetic average term is common to all genotypes in the chemostat):

$$\frac{1}{t}\left[\Lambda_i(t) \,-\, \Lambda_j(t)\right] \,=\, \frac{1}{t}\left[\Lambda_i^{\mathrm{rel}}(t) \,-\, \Lambda_j^{\mathrm{rel}}(t)\right] \tag{S14}$$

## S1.4 Polynomial modeling

We found that many of our normalized bar counts could not be accounted for by a model in which the growth rate per cell is constant over time. This would not be surprising if the thousands of interacting genotypes in the chemostat at once condition the growth medium beyond affecting the concentration of the common limiting nutrient, $S(t)$. If so, we expect a variety of time courses of $\lambda_i(t)$. One can think of these behaviors, in the context of the standard chemostat model [10], as an extension in which the maximum growth rate per cell, $\lambda_i^{\max}$, and the half-saturation constant, $K_i$ in the Monod growth rate function, are both functions of times, so that

$$\lambda_i(t) \,=\, \lambda_i^{\max}(t)\, \frac{S(t)}{K_i(t) \,+\, S(t)}. \tag{S15}$$

Our informal inspection of many normalized bar counts as a function of time suggested the possibility that the log of normalized counts could be fit by simple polynomial functions in $t$. Consequently, we modeled the exponential argument in Eq. (S9d), for the normalized count,

$$\gamma_i \,+\, \Lambda_i(t) \,-\, \mu_\Lambda^{\mathrm{a}}(t),$$

by a polynomial in $t$, up to 3$^{\mathrm{rd}}$ order; i.e,

$$\frac{\mu_i(t)}{s(t)} \,= \exp\left(\gamma_i \,+\, \Lambda_i(t) \,-\, \mu_\Lambda^{\mathrm{a}}(t)\right) \tag{S16}$$

$$\approx \exp\left(b_0 + b_1 t + b_2 t^2 + b_3 t^3\right) \tag{S17}$$

In the absence of terms higher than first order in the exponential argument in Eq. (S17), the coefficient $b_1$ is the fixed relative growth rate per cell in population $i$, $\lambda_i^{\mathrm{rel}}$.

We used DESeq2 to compute the $b$-coefficients in Eq (S17). Eq. (S17) is a standard generalized linear model with a log link function. Recall that we assume that the random barcode count $Y_i(t)$ is distributed as a negative binomial random variable with mean $\mu_i(t)$ and shape parameter $a_i$ (inverse of *dispersion*, as the term is used in DESeq2 [4]). We started our model fitting with the 3$^{\mathrm{rd}}$ order model and performed sequential model simplification by comparing the order-$k$ model to the order-$(k-1)$ model. For the comparison, we used the log ratio of maximum likelihoods test in DESeq2 [4], which gives a $p$-value and an adjusted $p$-value (False-Discovery Rate, FDR) for each simplification. All simplifications with an adjusted $p$-value greater than 0.1 were accepted in each round of simplifications. If $b_k$ was found to be significant, we did not test model simplifications at levels below $k$. The reason is that, if orthogonal polynomials [6] in $t$ were used as the basis for polynomial modeling, rather than powers of $t$, the orthogonal polynomial with highest order $k$, would, in general, include all powers less than $k$.

## S1.5 Fourier modeling for periodic environments

For our experiments with fluctuating environments, the chemostat input was switched periodically, in a square-wave manner, between two limiting nutrients, with period $T = 60\,\mathrm{hr}$. The corresponding frequency of the switching is $f = 1/T$. If we assume that the two limiting nutrients in the chemostat also fluctuate periodically at with same period, as confirmed in Fig. 1C, we expect to find some genotypes whose instantaneous growth rates per cell respond with periodic fluctuations, also with the same period. According to Fourier theory, a periodic function of $t$ for $0 < t < T$ can be expressed as a sum of sinusoidal components at integer multiples of the fundamental (input) frequency, $f = 1/T$.

We explored a model in which we assumed that the dynamics underlying $\lambda_i(t)$ are slow enough, so that $\lambda_i(t)$ is well approximated by a constant plus a single sinusoidal component at the fundamental frequency, $f = 1/T$. A mathematical formulation for this class of models is

$$\frac{1}{n}\frac{d}{dt}n = \lambda_0 + \lambda_1 \cos(2\pi f t + \theta), \qquad \text{or} \tag{S18a}$$

$$\frac{d}{dt}\log n = \lambda_0 + \lambda_1 \cos(2\pi f t + \theta), \tag{S18b}$$

where $\lambda_0$ and $\lambda_1$ are coefficients to be fit, and $\theta$ is the phase angle of the sinusoid. The solution to Eq. (S18b) is

$$\begin{aligned} \log\frac{n(t)}{n(0)} &= \lambda_0 t + \lambda_1 \frac{1}{2\pi f}[\sin(2\pi f t + \theta) - \sin\theta] \\ &= -\frac{\lambda_1}{2\pi f}\sin\theta + \lambda_0 t + \lambda_1 \frac{1}{2\pi f}\sin(2\pi f t + \theta) \end{aligned} \tag{S19}$$

Consequently we modeled the argument of the exponential in the equation for $\mu_i(t)/s(t)$ (expected value of normalized count) Eq. (S9d), for fluctuating environments, as

$$\gamma_i + \Lambda_i^{\mathrm{rel}}(t) \approx b_0 + b_1 t + b_2 \sin(2\pi f t) + b_3 \cos(2\pi f t). \tag{S20}$$

In Eq. (S20) we have used the trigonometric identity,

$$b_2 \sin(2\pi f t) + b_3 \cos(2\pi f t) = \sqrt{b_2^2 + b_3^2}\,\sin(2\pi f t + \theta),$$

where $\sqrt{b_2^2 + b_3^2}$ is the amplitude of the sinusoidal component of the response, and $\theta$ is the phase lag (angle) relative to the fundamental component of the periodic switching of the input limiting nutrient. The phase lag $\theta$ is given by $\arctan(b_2/b_3)$. We computed $\arctan(b_2/b_3)$ from the fitted parameters $b_2$ and $b_3$ by using the *atan2* function in R, which keeps track of the quadrant in which $\theta$ lies.

We tested the null hypothesis of no sinusoidal component (log maximum likelihood test), and we used the same FDR as that used in the polynomial modeling. We found roughly 700 out of roughly 4000 genotypes with significant sinusoidal fluctuation in the instantaneous growth rate per cell (minus the arithmetic mean). See Fig. 3E for an example of such a genotype.

Failure to reject the null hypothesis for a genotype made that genotype a candidate for a polynomial model. For sake of simplicity, we did not consider mixtures of $2^{\text{nd}}$ or $3^{\text{nd}}$ degree polynomials with an additional sinusoidal component.

## S1.6 Expectations for a chemostat with heterogeneous populations and fixed limiting-nutrient input: theory and computational modeling

### S1.6.1 Theoretical results on the possibility of fixed cell number and the coexistence of several genotypes with differing growth rate constants in the steady state

Our experimental results, with steady influx of a single limiting nutrient to the chemostat, show that, after a brief transient period before our experimental clock time $t = 0$, both the total cell number, $N(t)$, and the concentration of limiting nutrient, $S(t)$ (Figure 1D, left panel), appear to be constant. Meanwhile, the population proportions evolve, with decreasing Shannon diversity index (Figure 2B). At $t = 240\,\text{hr}$, a single genotype accounts for large proportion of total cells (Figure 2C). Because our experimental time period was only 240 hr (roughly 40 generations), one might inquire about the ultimate fate of cell number and population proportions, assuming that a steady state is achieved.

Although a straightforward extension of the basic chemostat model [10] for numerous heterogeneous populations (genotypes) in the chemostat [2] cannot reproduce the rich dynamical behavior we see in our experiments (S1.6.2), it is helpful to state the model equations here as a way of anchoring our thoughts. In the case of heterogeneous populations, the chemostat equations become,

$$\frac{d}{dt}N = \left[\mu_\lambda(t) - \beta\right] N \tag{S21a}$$

$$\frac{d}{dt}S = (S_0 - S)\,\beta - \frac{N}{Y}\mu_\lambda(t) \tag{S21b}$$

$$\lambda_i(t) = \lambda^{\max}\frac{S(t)}{K_i + S(t)}, \tag{S21c}$$

where: $N(t)$ is the total number of cells (per unit volume in this context); $S(t)$ is the concentration of limiting nutrient; $Y$ is the yield constant (see below); $K_i$ is the concentration of limiting nutrient that gives half the maximal growth rate constant for population $i$; $\lambda^{\max}$ is the maximal growth rate constant, assumed to be the same for all populations, for sake of simplicity; and $\mu_\lambda(t)$ is the population-mean growth rate per cell at time $t$, as defined in Eq. (S25) below.

It is worth noting that, in principle, a steady state for a chemostat with multiple populations and steady nutrient input to the chemostat is not guaranteed to exist. We show below that, if a true steady state is achieved, it is one in which only a single growth rate constant remains in the chemostat. In other words, the coexistence of multiple genotypes with different growth rate constants is not possible in a true steady state. Nevertheless, our simulations (S1.6.2) show that $N(t)$ and $S(t)$ can be very closely approximated by

constants along the way towards the steady state (S1.6.2), even as population proportions evolve towards an overall steady state in which all but one growth rate constant remains in the chemostat.

The argument for the ultimate steady state, if accessible, is based on evolution equations for total cell number, $N(t)$, genotype proportions, $p_i(t)$, and the population-mean growth rate, $\mu_\lambda(t)$. (Note absence of the superscript $a$, used previously to denote arithmetic mean, as opposed to the population mean.)

As a consequence of the fact that

$$\frac{d}{dt}\, n_i(t) \;=\; [\,\lambda_i(t) - \beta\,]\, n_i(t), \tag{S22}$$

the total number of cells in the chemostat,

$$N(t) \;=\; \sum_{i=1}^{m} n_i(t), \tag{S23}$$

evolves according to

$$\frac{d}{dt} N \;=\; \big[\,\mu_\lambda(t) \;-\; \beta\,\big]\; N. \tag{S24}$$

In Eq (S24), $\mu_\lambda(t)$ is the population-mean fitness. It is defined by

$$\mu_\lambda(t) \;=\; \sum_{i=1}^{m} p_i(t)\, \lambda_i(t), \tag{S25}$$

where $p_i(t)$ is the proportion of cells of genotype $i$ in the chemostat at time $t$; i.e.,

$$p_i(t) \;=\; \frac{n_i(t)}{N(t)}. \tag{S26}$$

The evolution equations for $p_i(t)$, are given by taking the derivative of both sides of Eq. (S26) (using the quotient rule and Eqs. (S22)–(S25)) to obtain

$$\frac{d}{dt}\, p_i(t) \;=\; \big[\lambda_i(t) \;-\; \mu_\lambda(t)\,\big]\; p_i(t). \tag{S27}$$

Eq (S27) says that, at every instant, $p_i'(t)$ is proportional to $p_i(t)$ with a proportionality factor that is the deviation of fitness $\lambda_i(t)$ from the population-mean fitness, $\mu_\lambda(t)$. The implication is that, for each genotype $i$, in the steady state, where $p,'\,(t) = 0$ (by definition of the steady state), either $\lambda_i = \mu_\lambda$, a single number, or $p_i = 0$. Consequently, in the steady state, there is either a single $\lambda_i \;=\; \mu_\lambda$ (by definition of the mean) with corresponding $p_i = 1$, or, there are several genotypes, all with the same $\lambda_i$ value, equal to $\mu_\lambda$ (again by definition of the mean), with a pooled proportion equal to 1. In either case, a single growth rate constant remains in the steady state (if a steady state is achieved).

In our experiments with a steady input rate of a single limiting nutrient, $N(t)$ appears to be constant, long before a steady state is achieved. This is inferred from our observation of substantial changes in population proportions throughout the time course of our experiments, despite the apparent constancy of $S(t)$ and $N(t)$. What insight can theory give us about this pre-steady-state regime?

According to Eq. (S24), $N'(t) = 0$, implies that $\mu_\lambda(t) = \beta$. By the same reasoning, if $N(t)$ is approximately constant, then $\mu_\lambda(t) \approx \beta$. This approximate equality can only be maintained in the long run if $\mu'_\lambda(t) \approx 0$.

This consideration motivates a look at the evolution equation for $\mu_\lambda(t)$ to discover what it is driven by. The rate of change of $\mu_\lambda(t)$ is given by taking the derivative of both sides of Eq. (S25) (using the product rule term by term), as follows:

$$\frac{d}{dt}\,\mu_\lambda(t) = \frac{d}{dt}\sum_{i=1}^{m} p_i(t)\lambda_i(t) \tag{S28a}$$

$$= \sum_{i=1}^{m} \left[p'_i(t)\lambda_i(t) + p_i(t)\lambda'_i(t)\right] \tag{S28b}$$

$$= \left[\sum_{i=1}^{m} \left[p'_i(t)\lambda_i(t)\right]\right] + \sum_{k=1}^{m} p_k(t)\lambda'_k(t) \tag{S28c}$$

$$= \left[\sum_{i=1}^{m} \left\{\left[\lambda_i(t) - \mu_\lambda(t)\right]\, p_i(t)\right\}\lambda_i(t)\right] + \sum_{k=1}^{m} p_k(t)\lambda'_k(t) \tag{S28d}$$

$$= \left[\left\{\sum_{i=1}^{m} p_i(t)\,\lambda_i^2(t)\right\} - \mu_\lambda^2(t)\right] + \sum_{k=1}^{m} p_k(t)\lambda'_k(t) \tag{S28e}$$

$$= \sigma_\lambda^2(t) + \sum_{k=1}^{m} p_k(t)\,\lambda'_k(t). \tag{S28f}$$

Eq (S28) says that the population-mean fitness at each time $t$ changes at a rate given by the instantaneous population-variance of fitness, $\sigma_\lambda^2(t)$ plus the population-mean rate of change of the fitness. Eq (S28) is a special case of the Price equation [7] in the continuous-time form [1], in which the numerical value of the fitness trait is fitness itself. See Eq. 2.4 of [1], with the phenotypic fitness trait quantified by their variable $z$ and their net reproductive rate $r$ both equal to our fitness $\lambda$, and with no instantaneous fitness difference $\Delta z$ between parent and offspring upon birth of the offspring. Note also that the covariance term in the Price equation, $\mathrm{cov}[z, r]$ in [1], in our case, becomes $\mathrm{cov}[\lambda, \lambda] = \mathrm{var}(\lambda)$.

From Eq. (S28) one can see that $\mu'_\lambda(t)$ can be small if both terms on the right-hand side of the equation are small, or if the terms counterbalance to give a small sum. Further insight is provided below by computational modeling.

According to Eq. (S28), if a steady state does exist (where, by definition of the steady state, $\mu'_\lambda(t) = 0$, and $\lambda'_i(t) = 0$ for all $i$), in this steady state,

$$\sigma_\lambda^2 = 0. \tag{S29}$$

The variance of fitness $\lambda$ in the steady state can only be equal to zero if all the $\lambda_k$ in a set $\mathcal{K}$ for which $p_k, \neq, 0$, are equal to one and the same value — call it $\lambda_\infty$. Here, we use $\lambda_\infty$ to mean $\lim_{t\to\infty} \lambda(t)$, which could

also be called $\lambda_{\rm ss}$, to refer to the steady state value. In other words, in the steady sate,

$$\lambda_k = \begin{cases} \lambda_\infty & \text{for } k \in \mathcal{K} \\ 0 & \text{otherwise} \end{cases} \tag{S30}$$

In our experimental protocol, it seems that, if there is more than one genotype with $\lambda_k = \lambda_\infty$, it is likely due to a number of gene knockouts that have no fitness phenotype.

Dean [2] discusses the practical circumstances in which multiple genotypes in a chemostat with different growth rates coexist with a fixed population number $N$. This regime corresponds to a situation in which $\mu_\lambda(t) - \beta \approx 0$, and, nevertheless, $p_i(t)$ change slowly to eventually give the true steady state in which only one growth rate constant per cell remains in the chemostat (illustrated below).

### S1.6.2 Computational results on the possibility of fixed cell number and the coexistence of several genotypes with differing growth rate constants in the steady state

We asked about the extent to which an extension of the standard mathematical model for a chemostat [10, 2], as given in the system of equations, Eqs. (S21), can account for our experimental findings in chemostats with steady input of a single limiting nutrient.

We simulated a chemostat with 4,000 populations, each with an independent random growth rate per cell. The results we report here are generic for heterogeneous populations and not dependent on a large number of different genotypes.

We found that, following a transient period of roughy 20 generation times (Fig. S1A), $S$ and $N$ eventually settled down to nearly steady values equal to those in the true overall steady state, $S_\infty$ and $N_\infty$, respectively (Fig. S1C). Note that the $y$-axis for relative $S$ and $N$ values in Fig. S1C spans only roughly 0.99–1.125. During a 200-generation time period, after the transient, with near constancy of $S$ and $N$, the population proportions evolve slowly (Fig. S1D), and the single dominant population achieves a proportion close to 1 at the end of this time period. Fig. S1B illustrates the dramatic change in the distribution of binned population proportions between the nominal end of the transient period of 20 generation times (Fig. S1B purple bars) and end of the simulation time, 220 generation times (Fig. S1B, grey bar). At the endpoint of the simulation time, only a single gray bar is visible, and the combined proportion genotypes it comprises are accounted for almost entirely by the winning population (consistent with Fig. S1D). In Fig. S1D the blue bars reflect the distribution of random $\lambda$-values at $t = 0$, but the $\lambda_i(S)$ values are evaluated t at $S = S_\infty$. The motivation is explained below.

**Parameter choices for the simulation** were made in the following way. The "winning" population is the one with the smallest value of the half-saturating constant, which we refer to as $K_1$. This single value of $K$ was chosen deterministically. To do so we first chose the steady-state value of the limiting nutrient

concentration, $S_\infty = 0.1$. We chose $K_1$ to enforce the constraint that $S_\infty = \varepsilon K_1$, where $\varepsilon << 1$; we chose $\varepsilon = 0.1$. In other words, the steady-state concentration of limiting nutrient $S_\infty$ is small compared to the half-saturating constant $K_1$. This is the typical setup of the chemostat model. We chose the single $\lambda^{\max}$ value, across populations, by satisfying the steady state condition, with only genotype 1 remaining in the chemostat. Namely that the growth rate constant is equal to the dilution rate constant; i.e., $\lambda_1(S_\infty) = \beta$. This is equivalent to

$$\lambda^{\max} \frac{S_\infty}{K_1 + S_\infty} = \beta, \tag{S31}$$

with the result that

$$\lambda^{\max} = \frac{1+\epsilon}{\epsilon}\beta. \tag{S32}$$

We chose the concentration of limiting nutrient in the input growth medium, $S_0$ to be large compared to the steady-sate value, $S_\infty$. In particular, we chose $S_0 = 10\, S_\infty$. This choice is also typical for the standard chemostat model. The steady-state density of cells in the chemostat, $N_\infty$, is determined by the yield constant $Y$ according to

$$\frac{N_\infty}{Y} = S_0 - S_\infty. \tag{S33}$$

Eq. (S33) reflects the fact that the chemostat model equations can be recast in terms of a normalized cell "density" given by $N/Y$. This means that either one can choose $N_\infty$, and this determines $Y$, or one can choose $Y$, and this determines $N_\infty$. We adopted the former approach and set $N_\infty$ to an arbitrary, but reasonable value of $2.5 \times 10^7$/mL, as in Figure 3 of [10]. The distribution of the random $\lambda_i$ values at any value of $S$ is determined, with fixed $\lambda^{\max}$, by the distribution of $K_i$ values, or vice versa. We know that the largest $\lambda(t)$, as $t \to \infty$ is

$$\lambda_1(S_\infty) = \lambda^{\max} \frac{S_\infty}{K_1 + S_\infty} = \beta. \tag{S34}$$

And we know that the other $\lambda_i(S_\infty)$ values are less than $\lambda_1(S_\infty)$. This motivates us to choose $\lambda_i(S_\infty)$ from a distribution with maximum value equal to $\beta$, and a minimum value considerably lower; we took this value to be equal to $0.1\,\beta$. We used a truncated gamma density function for sake of convenience, with a shape parameter equal to 16 (giving a standard deviation of 0.25) and a mean value equal to $\beta/2$. For any randomly chosen $\lambda_i(S_\infty)$ value, the corresponding $K_i$ value is given by

$$K_i = \left( \frac{\lambda^{\max}}{\lambda_i(S_\infty)} - 1 \right) S_\infty. \tag{S35}$$

# References


[1] Troy Day, Todd Parsons, Amaury Lambert, and Sylvain Gandon. The price equation and evolutionary epidemiology. *Philos Trans R Soc Lond B Biol Sci*, 375(1797):20190357, 04 2020.

[2] Antony M Dean. Protecting haploid polymorphisms in temporally variable environments. *Genetics*, 169(2):1147–56, Feb 2005.

[3] Daniela Delneri. Barcode technology in yeast: application to pharmacogenomics. *FEMS Yeast Res*, 10(8):1083–9, Dec 2010.

[4] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol*, 15(12):550, 2014.

[5] Davis J McCarthy, Yunshun Chen, and Gordon K Smyth. Differential expression analysis of multifactor rna-seq experiments with respect to biological variation. *Nucleic Acids Res*, 40(10):4288–97, May 2012.

[6] Sabhash C. Narula. Orthogonal polynomial regression. *International Statistical Review / Revue Internationale de Statistique*, 47(1):31–36, 1979.

[7] G R Price. Selection and covariance. *Nature*, 227(5257):520–1, Aug 1970.

[8] Matthew E Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res*, 43(7):e47, Apr 2015.

[9] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–40, Jan 2010.

[10] Naomi Ziv, Nathan J Brandt, and David Gresham. The use of chemostats in microbial systems biology. *J Vis Exp*, (80), Oct 2013.

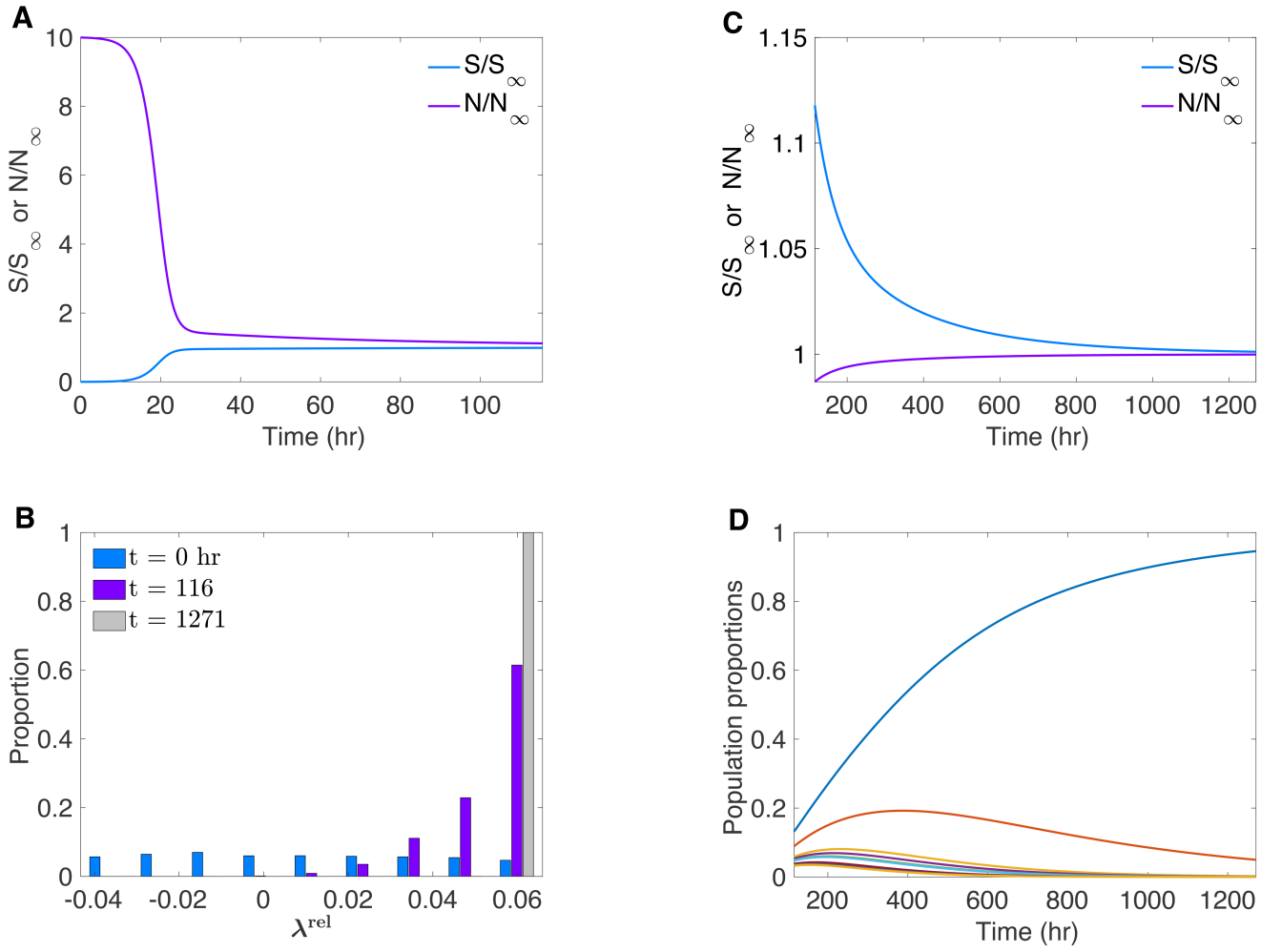


Figure S1: **Population dynamics in a chemostat with heterogeneous genotype fitness and static input of limiting-nutrient growth medium** (A) Evolution of $N(t)$ and $S(t)$ towards steady state values starting from initial conditions $N(0) << N_\infty$, and $S(0) = S_0$, the concentration of limiting nutrient in the chemostat input medium. Note that the variable are expressed in relative terms $N(t)/N_\infty$, and $S(t)/S_\infty$, so that, as $t \to \infty$, both relative values approach 1. (B) Distribution of $\lambda^{\text{rel}}$ values (binned proportions) just after an initial transient time of 20 generation times (purple bars) and at 200 generation times thereafter (gray bars). Blue bars reflect distribution of $\lambda$ values at $t = 0$, as explained in the text. (C) Relative $N(t)$ and $S(t)$, showing slow approach towards a value of 1 with very small deviations over a 200-generation time period following an initial transient period. (D) Slow evolution of population proportions (top 10), while $N(t)$ and $S(t)$ hardly deviate form their steady-state values.