Scalable continuous culture-based approaches for understanding evolution and cancer

Aaron W. Miller

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Maitreya Dunham, Chair

Stan Fields

Eric Klavins

Program Authorized to Offer Degree:

Genome Sciences

University of Washington

Abstract

Scalable continuous culture-based approaches for understanding evolution and cancer

Aaron W. Miller

Chair of the Supervisory Committee:
Maitreya Dunham
Associate Professor of Genome Sciences
Department of Genome Sciences

Understanding the genetic basis of adaptation in evolution is an area of research that has the potential to characterize the forces that have shaped the evolution of all life. For much of history, it was far easier to see evolution's effect than its genetic cause; however, modern sequencing methods are for the first time making it possible to comprehensively characterize the genetic basis of evolution. As such, there is a renaissance for experimental microbial evolution, as researchers attempt to characterize the genetic basis of adaptation in a given environment. However, controlling the environment is an ongoing challenge in the field. Use of continuous culture is an ideal way to control the environment in which microbes are grown and to provide an invariant selection over time. One disadvantage of continuous cultures is that they are generally carried out in low throughput, thus limiting what we can learn from them. The focus of

my dissertation has been the marriage of multiplexed continuous cultures that I developed in graduate school with selection and sequencing techniques to allow for a more comprehensive view of the genetic basis of adaptation in *Saccharomyces cerevisiae*. Additionally, I have developed a new continuous culture-based method for multiplexed determination of mutation rate, which I have applied to the study of mismatch repair genes relevant to Lynch syndrome risk in humans.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank Maitreya for providing an excellent environment for me to grow as a scientist these past six years. I have learned a great deal and had a lot of fun while working for you and for that I will be forever thankful!

I would like to thank my labmates for being excellent friends and coworkers and extended family for my time in graduate school – you've been great! Additionally I am grateful for having a great cohort that I hope to remain friends with forever. I feel grateful for the education I have received in the Genome Sciences Department and am truly humbled and excited by the quality of work continually being produced by the friendly people here.

I would like to thank my committee for their support and wisdom. I have grown in graduate school and they have helped my growth a great deal. Thank you!

Finally none of this would be possible without the support of my family. Thanks to my Mother, Father, and Sister for being there to support me and have fun with. Additionally I would like to thank Rosie for everything! You mean the world to me!!!

Chapter 1. INTRODUCTION

1.1 WHY STUDY MICROBIAL EVOLUTION OR EVOLUTION AT ALL?

Natural selection is a cornerstone of modern biology and is fundamental to our understanding of how all life arose on earth. Natural selection was originally described in 1858 in works by Darwin and Alfred Russel Wallace and shortly thereafter in Darwin's *On the origin of species*, which proposed that natural selection is analogous to artificial selection incurred by breeding of animals and plants with desirable traits by humans. Adaptation is the result of selection. Adaptation is characterized by the movement of a population towards a phenotype that best fits the present environment. Natural selection acts on any heritable phenotypic trait, and selective pressure can be produced by any aspect of the environment, including sexual selection and competition with members of the same or other species. Traits that cause greater reproductive success of an organism are said to be selected for, whereas those that reduce success are selected against. Until very recently, those who would like to understand evolution were left to draw conclusions from the effect of evolution through observation of phenotypes without knowing the underlying genetic cause. As such, many important questions about the genetic basis of adaptation remain unanswered by evolutionary theory and are central questions for experimental evolution. What types and specific mutations are responsible for adaptation? Do most adaptations involve single genes of large phenotypic effect or many small ones? How repeatable are evolutionary trajectories and what does recurrence of mutations within a given gene or pathway tell us about the fitness landscape of the evolving organism? How do parameters such as population size, strength of selection, and genetic background influence rates and mutations associated with adaptation? Can we describe the distribution of phenotypic effects

among the mutations that are observed during adaptation? These basic questions that underlie evolution are not answered by traditional evolutionary theory and rely on a sufficiency of experimental data.

Experimental microbial evolution generally entails selection of *de novo* mutations that arise during prolonged growth of initially genetically clonal populations. Experimental evolution in microbes differs from experimental evolution in animals such as worms (Teotonio, Chelo et al. 2009), flies (Turner and Miller 2012), and mice (Swallow, Koteja et al. 1999) which typically entails selection on pre-existing (standing) variation genetic variation through reproduction between heterogeneous individuals present in the founding populations. Use of microbes comes with a number of distinct advantages over other systems; for example, it is common to maintain large population sizes (10^8-10^{10}) of individuals with short generations times (30-360 minutes) that typically have small genome sizes (10^6-10^7 bases). In both experimental and natural populations, mutations responsible for adaptation include single nucleotide changes ((Steiner, Weber et al. 2007), (Barrick and Lenski 2009)), transposition events ((Wilke and Adams 1992), (Aminetzach, Macpherson et al. 2005)), and copy-number variants including gene amplifications and deletions ((Brown, Todd et al. 1998), (Perry, Dominy et al. 2007)). The importance of each mutational class is not known and varies widely based on the selection imposed. For *Saccharomyces cerevisiae* the rate of point mutation is 10^{-9}–10^{-10} per base pair/generation (Lang and Murray 2008), whereas retrotransposition is estimated to occur at roughly 10^{-8} events per generation (Qian, Huang et al. 1998). Mutations that confer a loss of gene function are estimated to occur as often as 10^{-5} events per generation (Thomas and Rothstein 1989). Finally, the spontaneous rate of gene amplification events has been estimated to occur as infrequently as 10^{-10} events per generation (Dorsey, Peterson et al. 1992). Thus spontaneous mutation rates for

many experimental evolution scenarios are thought to on average generate every possible single base substitution in the genome in the population each generation. Additionally many deletions, rearrangements, and amplifications add ample variation upon which to select.

One final advantage of using microbes to study experimental evolution is the potential to record a frozen fossil record throughout an experiment, from which one can return to any population at any time-point collected in an experiment and measure fitness, assess phenotype, identify mutations or pick up where the experiment left off. Since experimental evolution is often used on traditional simple model systems such as phage, bacteria, and yeast, a wealth of genetic tools are available to aid in the downstream genetic analysis of evolved populations. Together, these many advantages make the study of experimental microbial evolution an ideal tool for characterizing the molecular basis of adaptation and answering basic questions about how evolution transpires.

1.2 Advantages and limitations of continuous culture in experimental evolution

Long-term growth can be achieved through a variety of methods. Serial batch transfer (see **Fig 1.1A.)** is the most common, based on ease of implementation. The low cost, ease of culture propagation, and scalability through use of 96-well plates and culture flasks have all contributed to the prevalence of this culture method. However, in batch culture, the chemical environment is continually changing as cells grow, divide, consume nutrients, and excrete waste products. These changes associated with growth have substantial effects on cellular physiology (Saldanha, Brauer et al. 2004). Continuous cultures, chemostats and turbidostats, stand out as an attractive alternative of increasing popularity (Hoskisson and Hobbs 2005) as they are the only way to impose a continuous and invariant selective pressure in cultured cells.

A chemostat is simply a culture of constant volume in which for each drop of media added a drop of culture is removed (see **Fig 1.1B.**). Cells growing in a chemostat quickly reach a steady state in which the chemical environment, composed of nutrients available and molecules secreted, becomes static (hence the name "chemo-stat"). An essential requirement of the chemostat is the use of a defined medium in which a single nutrient is present at a growth limiting concentration (Ziv, Brandt et al. 2013). A nutrient is said to be limiting when its concentration determines steady-state cell density; all other required nutrients are present in excess. During steady-state growth the chemical environment and selection imposed is invariant and can be maintained indefinitely. This consistent selection is ideal for experimental evolution and allows for a simpler association between adaptive mutations and the environment in which they arose. Studies of microbial physiology in chemostats have made two important observations about growth at steady-state (Saldanha, Brauer et al. 2004): 1) that cells experience a consistent level of gene expression across replicate chemostat cultures of the same nutrient-limitation; and 2) that cells experience a state of hunger but not starvation similar to that experienced by cells approaching but not in stationary phase in a batch culture. One important criterion for what nutrients can be used for chemostat growth is that they must be essential for the growth of the organism. Carbon, nitrogen, phosphorus, and sulfur sources are most common. Non-essential nutrients can be made essential through the use of auxotrophic strains that then may be used in chemostats. Improved fitness in the chemostat environment is typically achieved through improved ability to acquire or utilize the growth-limiting nutrient and subsequently divide. The fitness of the evolving population can be measured through direct competition against a genetically marked strain (see **Fig 1.2.A.**).

Turbidostats are similar to a chemostat in that the culture is continuously diluted with fresh medium and the culture volume is held constant through continual removal of an equivalent volume of culture. Where turbidostats differ is that they use a feedback control to keep cell density constant; in a turbidostat, the cell density is continually monitored and used to compute the dilution rate that achieves a prespecified cell density (see **Fig 1.1C.**). As such, cells grown in a turbidostat below maximal turbidity avoid ever experiencing nutrient limitation. The resultant steady-state environment is most similar to batch culture during the mid-log exponential phase of growth, when growth is maximal and nutrients are in abundant supply. Thus growth rate is determined by the intrinsic properties of the cell that determine the rate of cellular replication. In this way the rate of cell division and thus fitness of a culture is directly proportional to the dilution rate. (**See Fig 1.2.B.**)

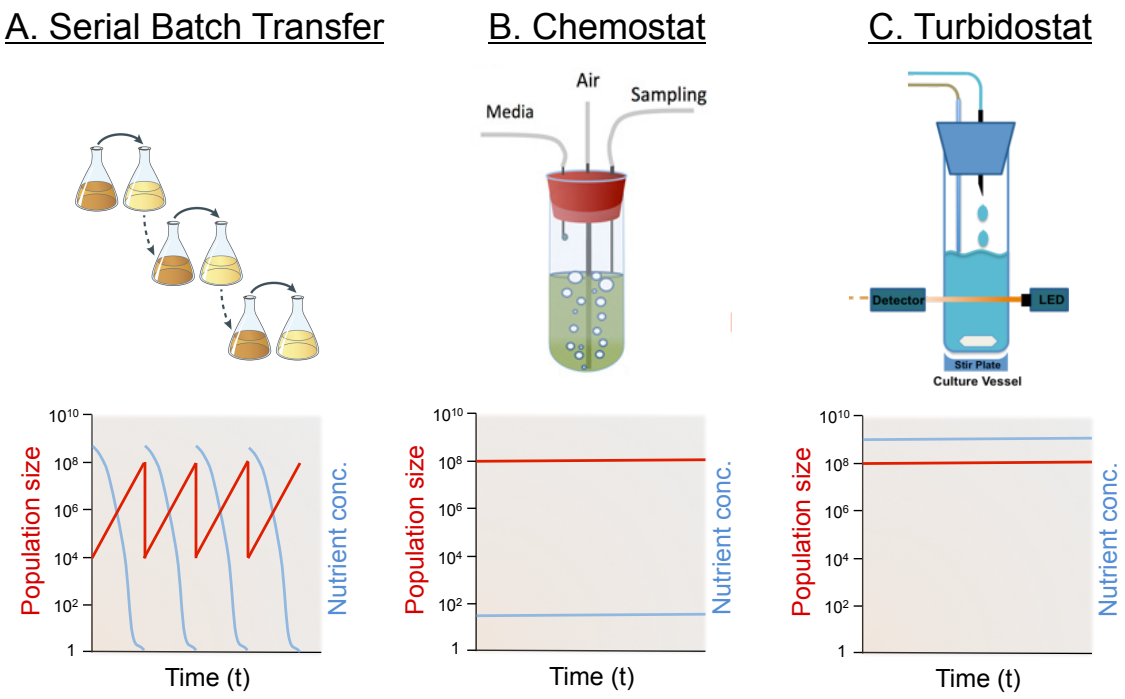


Figure 1.2A, B, C. Three main ways in which populations are propagated in evolution experiments that all lead to different types of genetic dynamics. **A.** In serial transfer experiments a proportion of the population is passaged to fresh media and allowed to regrow, most typically until the limiting nutrient is exhausted. Batch grown cultures experience variations in nutrient level and cell density over each growth cycle. **B.** In chemostats populations are maintained in conditions of constant inflow of nutrients and outflow of culture and waste where population size is constrained by nutrient availability of a specific limiting nutrient that is kept low. **C.** In turbidostats populations grow at their maximal growth rate by tuning dilution rate to maintain a desired culture density.

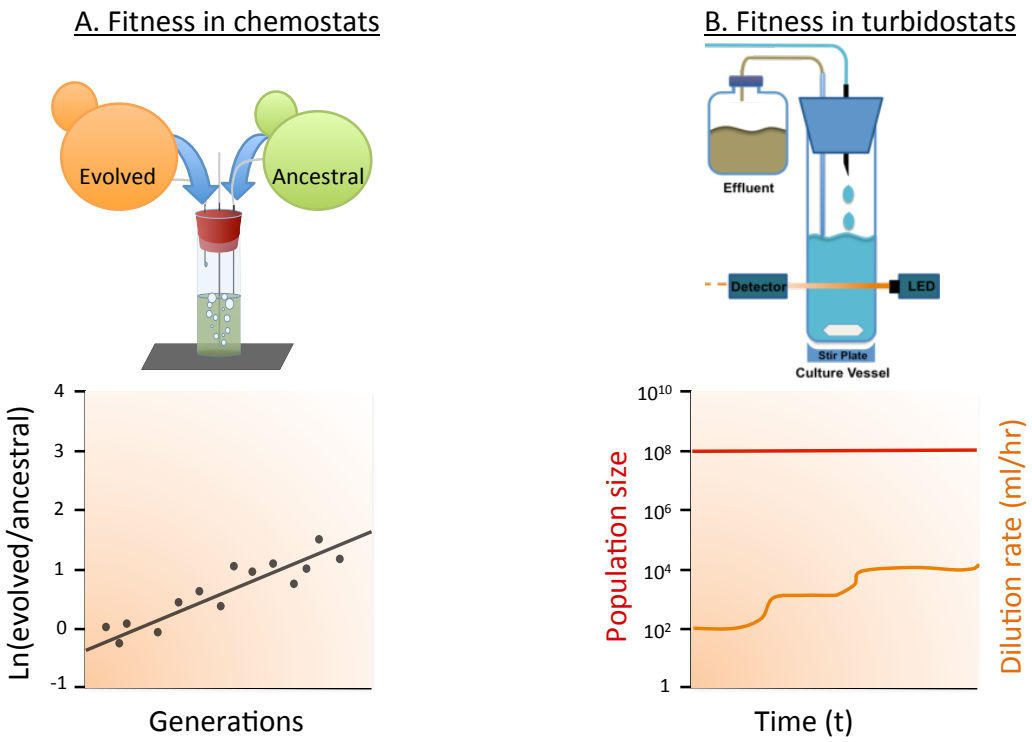


Figure 1.2.A,B. How fitness of an evolving culture is measured. A) In chemostats the fitness of a culture is generally determined by direct competition versus a genetically marked strain. The proportion of one relative to the other over time is used to derive a relative fitness measurement known as the selection coefficient. B) In turbidostats fitness is directly proportional to the dilution rate.

One important limitation of both chemostats and turbidostats is that both also select for cells that clump and adhere to the vessel walls. Mutations that cause cell clumping and thereby prevent these cells from being removed from the chemostat thus bypass the selection imposed by nutrient limitation. Acquisition of these mutations generally marks the end of a continuous culture experiment and thus limits ultra long-term evolution experiments. In this way obtaining a starting genetic background wherein clumping occurs less frequently is a common desire of the microbial experimental evolution community. Despite this limitation, the ability to control and simplify selection the way that continuous cultures allow profoundly improves our ability to answer fundamental questions about how adaptation proceeds in a given environment.

1.3 GENOME SEQUENCING REVOLUTION AND THE COMPLEXITY OF ADAPTATION AT THE LEVEL OF WHOLE GENOMES.

Although experimental evolution with microbes commenced 130 years ago with the work of Darwin's contemporary, Reverend W.H. Dallinger, until recently the field has been limited by the inability to comprehensively characterize the genetic variation associated with adaptive evolution. Prior to the development of next-generation sequencing, Sanger sequencing could be successfully used to characterize adaptation at the whole genome level for phage that have genomes on the order of a few kilobases. Adaptive evolution for more complex organisms was greatly restricted in terms of scope and could be characterized only on a gene-by-gene basis. While this candidate approach did find some mutations responsible for adaptation, it was completely insufficient to comprehensively identify mutations across the genome.

Genotyping technologies, in particular next-generation sequencing, have removed this limitation and ushered in a new phase of experimental evolution where we now can for the first

time easily characterize mutations in evolving clones and populations. Some of the first evolved clones to be sequenced were in Richard Lenski's lab, where they sequenced single clones sampled from five time-points across 20,000 generations (Barrick, Yu et al. 2009). In this study, Barrick *et al*. (2009) showed that mutations continue to accumulate steadily through time despite a decrease in the rate at which fitness increases. This early and well-characterized example of 'diminishing returns' is rooted in the discrepancy between the rate at which mutations arise and their effect on fitness, which hints at the importance of epistasis in shaping adaptation.

A wide variety of mutations have been identified in *S. cerevisiae* grown in glucose, nitrogen, phosphate, and sulfate-limited chemostats. One of the most commonly reported classes of mutations is amplification of specific nutrient transporters relevant to the environment the cells are grown in. For example, amplification of the glucose, amino acid, phosphate, and sulfate transporters (*HXT6/7*, *GAP1*, *PHO84*, and *SUL1*, respectively) were detected in clones (Brown, Todd et al. 1998), (Gresham, Usaite et al. 2010), (Levy, Kafri et al. 2011), (Gresham, Desai et al. 2008). Multiple evolution experiments from near identical ancestral populations resulting in mutations in the same gene or in genes with similar function highlight an important finding of the experimental evolution community - that cases of parallel evolution are common. Reoccurrence of mutations within a gene or pathway across an evolution reveals functions important for adaptation. Furthermore, recurrence can be used to highlight what mutations are likely to be adaptive rather than incidental. While it is not always apparent what the significance of a particular class of mutation may be and how its frequency relative to others influences adaptation, generally these mutations fall into two classes: 'Loss Of Function' or 'LOF' mutations and 'Gain Of Function' or 'GOF' mutations. In this paradigm mutations that truncate or cause a frame-shift in a protein-coding region, especially near the 5' end, are interpreted as

likely causing a LOF, whereas amplifications, especially of nutrient transporters like those described above, have generally been shown to result in a GOF with regard to improved ability to import a specific limiting nutrient and a concordant increase in relative fitness. Missense mutations and synonymous mutations are somewhat more difficult to interpret and their importance in a specific context may be tested directly through allele replacement experiments or inferred through algorithms that rely on conservation or other parameters that attempt to reveal the particular importance of a given amino acid in a protein. Indeed, determining the effect of specific mutations that arise in evolution experiments is an ongoing challenge in the field.

Even when individual mutations have a reasonably straightforward interpretation, it is much less clear how these mutations interact in the complex genotypes typical of evolved strains. Further analysis of clones derived from evolving microbial populations has provided specific examples of how epistasis can constrain and guide evolution (Kvitek and Sherlock 2011), (Barroso-Batista, Sousa et al. 2014). One example of this is the observation that haploid yeast evolved in glucose-limited media in chemostats repeatedly obtain loss of function mutations in *MTH1* (negative regulator of the glucose-sensing signal transduction pathway) or an amplification of the hexose transporters *HXT6,7* but never both (Kvitek and Sherlock 2011). Individually either mutation results in increased fitness (~20% for alleles of *mth1* and ~30% for amplifications of *HXT6,7*); however, the fitness of a strain engineered to have both mutations (*mth1* and *HXT6,7*) had a fitness deficit of 10% below that of the ancestral strain. When two beneficial mutations have a combined net-negative effect in a particular environment, this is known as reciprocal sign epistasis, which is one intuitive example of how epistasis can constrain which beneficial mutations co-occur and thus shape adaptive trajectories.

To further complicate matters, multiple genotypes may be present in a single evolving population. Deep sequencing of evolving populations has provided additional insights into how microbial asexual populations evolve both in terms of mutations that arise and with regard to their prevalence in the population. Previously, it was not clear whether beneficial mutations would more often arise in succession to sweep the population or if different subpopulations would more frequently stably coexist, left to compete for limiting nutrients across hundreds of generations. Two recent studies that employed population sequencing clearly show pervasive clonal interference and long-term coexistence of genetically distinct subpopulations that compete for limiting nutrients and prevent each other from completely sweeping the population (Kvitek and Sherlock 2013, Lang, Rice et al. 2013). In these early examples of whole population sequencing, the aforementioned case of reciprocal sign epistasis between *mth1* and amplification of *HXT6,7* are seen independently within subpopulations of the same evolving population. This observation raises the question of whether epistasis not only influences adaptive trajectories but also promotes evolutionary divergence resulting in clonal interference (Kvitek and Sherlock 2013, Lang, Rice et al. 2013).

In summary, microbial experimental evolution has demonstrated that adaptation can occur through a variety of mechanisms including amplifications, insertions or deletions, single nucleotide variants and large-scale structural variation. A given selection regime, particularly in consistent selections like those seen in chemostats, often repeatedly selects for mutations in the same gene, or genes with similar function. This 'convergent' evolution is seen across a wide range of evolution experiments, and is in many ways indicative of a simple fitness landscape. Where evolution experiments yield different results across replicate evolution experiments, it is thought that this may be indicative of 'rugged' landscape. This assertion is in fact common in

experimental evolution literature and is a reasonable conjecture; however, it may simply be that more biological replicates are necessary to reveal convergent adaptive strategies. Microbial experimental evolution research is also beginning to provide specific examples of how greatly evolution is influenced by epistasis. The identity and order in which mutations are acquired, if they co-occur, and how many subpopulations exist in an evolving population can be, in part, attributed to the effects of epistasis. However, most studies have few biological replicates per genotype for a particular organism. Clearly, the field could benefit from more experimental replication to better quantitate how common and impactful epistasis is in shaping evolutionary trajectories for a particular species in a particular selective regime. However, no continuous culture technology existed to do such a thing at the required scale. In order to tackle these and other evolutionary questions, I first had to build the platform on which to do so. The design and validation of the multiplexed chemostat arrays (**Chapter 2**) enable the user to carry out tens or hundreds of evolution experiments at once. This platform should facilitate more comprehensive characterization of evolutionary trajectories for more thorough answers to ongoing evolutionary questions.

1.4 COMMON METHODS TO CHARACTERIZE MUTATION RATE

Mutation rate is an important parameter in evolution, for without mutations there would be no variation upon which to select. As such, the mutation rate impacts the speed of adaptation in a given environment. However, mutation rates must be balanced: too high and a population will acquire excess deleterious mutations with each cell division and will not thrive. Too low of a mutation rate will result in slower adaptation, a great disadvantage for an organism growing as part of an ecological community in nature where the environment and selective pressures are far

from constant. Mutation rates vary between species, between individuals of the same species, within regions of the genome, and between environments. The importance of mutation rate is particularly evident in the struggle between pathogenic strains and the host immune system, and in the evolution of cancer, where in both cases an elevated mutation rate imbues these cell types with a competitive edge essential and intrinsic to their pathology.

Despite the importance of mutation rates, actual quantitation of mutation rate has not been performed for almost all species. Two common experimental methods used to measure mutation rate in single cells include the mutation accumulation assay in chemostats and the fluctuation assay. The fluctuation assay tracks the accumulation of resistance markers across replicate batch cultures (Luria and Delbruck 1943). To account for the jackpot effect created by early resistance mutations, this assay quantifies the number of replicate cultures that produce no resistant cells to calculate the mutation rate. As such, the fluctuation assay requires many replicates and can only be used to measure mutation rate for one allele at a time.

Chemostats are an excellent tool to quantify mutation rate (Novick and Szilard 1950) as they are free from the extreme fluctuations of batch cultures described by Luria and Delbruck (Luria and Delbruck 1943). During each doubling of the chemostat culture, half of the cells are diluted out, leading to a linear accumulation of neutral mutations at the mutation rate. Therefore the mutation rate can be directly measured by calculating the slope of the mutation frequency over many generations of growth. To date, there is still no general multiplexed system to quantitate mutation rate of many species, strains, or alleles that confer variance in mutation rate. The study of evolution, pathogenicity, and cancer risk would benefit from a high-throughput way to accurately quantitate mutation rate. To solve this problem I have developed a promising scalable means to measure mutation rate for many genotypes at once by exploiting the benefits

of continuous culture and next generation sequencing. Thus far, I have prototyped this method in yeast on alleles of mismatch repair alleles identified in Lynch Syndrome patients (**Chapter 5**) although the described methods are general and could likely be used in a wide variety of contexts and cell types.

1.5 DISSERTATION OBJECTIVES

To address the lack of biological replication in experimental evolution and general lack of means to functionally characterize mutations relevant to evolution and cancer, in collaboration with members of the Klavins, Thevelein, and Gammie Labs I have:

- Developed and validated a multiplexed miniature chemostat array (**Chapter 2**).
- With a collaborator from the Klavins lab worked to develop flexible open-source turbidostats based in part on my miniature chemostat array designs. (**Appendix A**).
- I have subsequently used the chemostat array for a multitude of experiments, some of which are described in my studies of experimental evolution in (**Chapter 3**). This work included 300 generations of evolution for 95 cultures, during which competitions were performed every 50 generations to measure rates of adaptation. I performed whole genome whole population sequencing to survey occurrence and reoccurrence of specific mutations across 3 nutrient environments: glucose, phosphate, and sulfate limitation.
- The list of mutations observed in the evolution experiments in (**Chapter 3**) were compared to many other relevant evolution experiments present in the experimental evolution literature presented in (**Appendix B**).
- To better understand how genetic background influences evolutionary trajectories I performed compensatory evolution experiments detailed in (**Chapter 4**), for which both the *sul1*Δ background and *sul1*Δ *sul2*Δ background was grown for prolonged periods of

time in sulfate-limited media. In doing so I discovered a novel low-affinity sulfate transporter that can be easily mutated into a high-affinity sulfate transporter.

- I have developed a promising scalable means to measure mutation rate for many genotypes at once by exploiting the benefits of continuous culture and next generation sequencing (described in **Chapter 5**). I have also initiated experiments to allow for multiplexed characterization of mutation rate amongst multiple genes involved in mismatch repair.

Chapter 2. DESIGN AND USE OF MULTIPLEXED CHEMOSTAT ARRAYS

2.1 ACKNOWLEDGEMENTS

In this acknowledgements section, and those that follow, I detail my own contributions as well as the significant contributions of collaborators.

I developed the miniature array of chemostats 'ministats' described in this chapter primarily during my rotation in the Dunham lab and validated them shortly after joining the lab in my second year. I generated, analyzed, and made figures for all data presented in this chapter with the exception of the expression data that were generated with help from Emily Mitchell and Maitreya (**Fig 2.2.B.**). The contents of this chapter were published in JoVE on February 23, 2013 along with a video entailing how to make and use ministats filmed and edited by Corrie Befort.

Miller AW, Befort C, Kerr EO, Dunham MJ. Design and Use of Multiplexed Chemostat Arrays. *Journal of Visualized Experiments : JoVE*. 2013;(72):50262. doi:10.3791/50262.

2.2 ABSTRACT

Chemostats are continuous culture systems in which cells are grown in a tightly controlled, chemically constant environment where culture density is constrained by limiting specific nutrients (Monod 1950, Novick and Szilard 1950). Data from chemostats are highly reproducible for the measurement of quantitative phenotypes as chemostats provide a constant growth rate and environment at steady state. For these reasons, chemostats have become useful tools for fine-scale characterization of physiology through analysis of gene expression(ter Linde, Liang et al. 1999, Boer, de Winde et al. 2003, Wu, Zhang et al. 2004, Daran-Lapujade, Daran et al. 2009) and other characteristics of cultures at steady-state equilibrium (Diderich, Schepper et al. 1999). Long-term experiments in chemostats can highlight specific trajectories that microbial populations adopt during adaptive evolution in a controlled environment. In fact, chemostats have been used for experimental evolution since their invention (Novick and Szilard 1950). A common result in evolution experiments is for each biological replicate to acquire a unique repertoire of mutations (Kubitschek and Bendigkeit 1964, Helling, Vargas et al. 1987, Gresham, Desai et al. 2008, Kao and Sherlock 2008, Kvitek and Sherlock 2011). This diversity suggests that there is much left to be discovered by performing evolution experiments with far greater throughput.

2.3 INTRODUCTION

We present here the design and operation of a relatively simple, low cost array of miniature chemostats - or ministats - and validate their use in determination of physiology and in evolution experiments with yeast. This approach entails growth of tens of chemostats run off a single multiplexed peristaltic pump. The cultures are maintained at a 20 ml working volume,

which is practical for a variety of applications. It is our hope that increasing throughput, decreasing expense, and providing detailed building and operation instructions may also motivate research and industrial application of this design as a general platform for functionally characterizing large numbers of strains, species, and growth parameters, as well as genetic or drug libraries.

The dynamics of microbial growth and evolution are fundamental to microbiology, ecology, genetics, and biotechnology. The most common method of culturing microbes is in batch, where cells are inoculated at low density into nutrient-rich broth and grown to saturation. Though straightforward to perform using standard laboratory equipment, batch cultures experience a fluctuating chemical environment and correspondingly changing cellular physiology. This heterogeneous growth environment can result in secondary growth and stress effects that can mask subtle physiological differences. Experimental evolution by serial batch transfer can select for complex mixtures of growth-phase specific subpopulations, complicating attempts to connect adaptations to specific selective conditions. Measurement of quantitative phenotypes can be difficult due to noise from imprecise sample timing and variation in features such as lag time. Continuous cultures provide an alternative growth regime where cells can be reproducibly cultivated in a chemically homogeneous environment at a defined growth rate to reach a physiological steady state. Because of these advantages, studies of experimental evolution and characterization of cellular state often utilize the controlled environment of continuous cultures like the chemostat (Dunham 2010).

Appreciation of these advantages has led to a resurgence in interest in chemostat cultures (Hoskisson and Hobbs 2005). Since their introduction in 1950 (Monod 1950), (Novick and Szilard 1950) chemostat systems have been developed to function on a variety of scales ranging

from liters to microliters and for a variety of applications (Balagadde, You et al. 2005, Groisman, Lobo et al. 2005, Nanchen, Schicker et al. 2006, Klein, Schneider et al. 2013). These various designs, which range from commercially produced bioreactors to glass-blown vessels to custom microfluidics platforms, share general design principles. A culture chamber is stirred and aerated (usually by bubbling air through it) and the microbes contained therein are kept homogeneously dispersed throughout the culture chamber at all times. Fresh medium of defined composition is added continually and the rate of addition controls growth rate and influences the chemical environment experienced by the culture. An overflow sets the culture volume in the growth tube, and through this overflow the culture will be sampled at the same rate at which fresh medium enters. In this way cultures quickly reach a physiological steady state at which many biological parameters remain constant. Despite the advantages of chemostats and reports of these various platforms in the literature, widespread adoption has been limited by difficulties in building and operating these systems, and high costs associated with commercial options. Furthermore descriptions of how to make and use these devices can be opaque.

We observe highly consistent experimental parameters and reproducible results when comparing our device to reported data for yeast cultured in larger volume commercial bioreactors. For example, we observe reproducibility of cellular physiology as seen through reaching steady-state equilibrium within 10-15 generations and obtaining similar culture densities at equilibrium. Additionally, gene expression patterns are consistent between ministats and a commercial larger-volume platform. Stability of dilution rate, optical density and reproducibility of gene expression between three replicate cultures demonstrate the robustness of our platform. We also show that the same adaptive mutations arise over similar experimental evolution timescales as with larger volume chemostats.

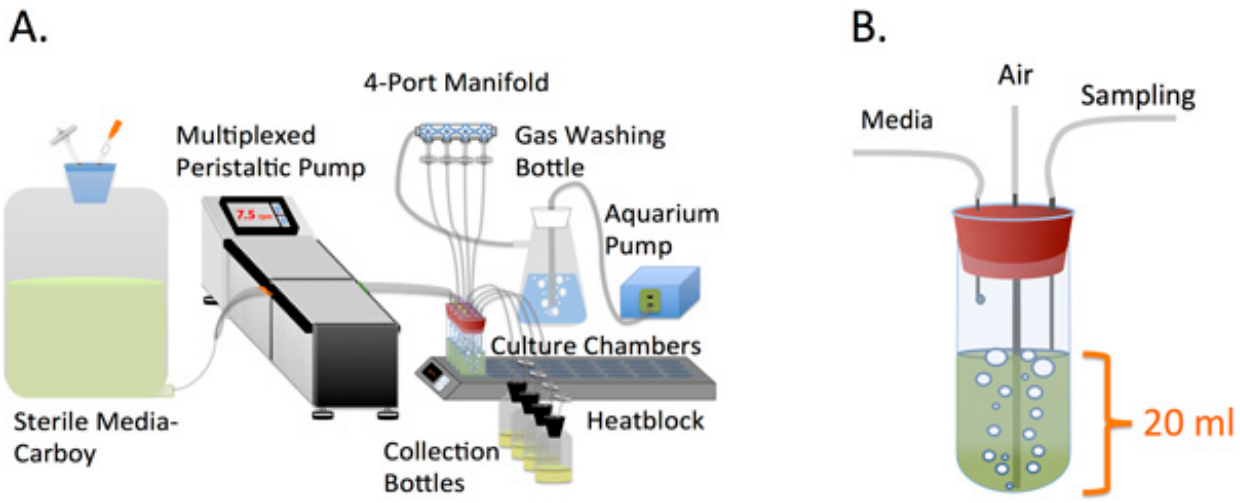


Figure 2.1A,B. A. Design and arrangement of the ministat array. **B.** Design of the culture chamber.

2.4 RESULTS

The ministat array described above and in (**Fig 2.1A, B.**) was used to culture a haploid *MATa* laboratory strain of budding yeast (S288c) under sulfate-limiting conditions as described previously (Gresham, Desai et al. 2008). We tested efficacy for common chemostat applications including determination of physiology and experimental evolution. To validate the ministats, we repeated several experiments previously performed in industrial fermentors modified for chemostat use (Brauer, Saldanha et al. 2005, Torres, Sokolsky et al. 2007, Gresham, Desai et al. 2008). ATR Sixfors fermentors were run at a 300 ml working volume, over ten times the volume of the ministats, and have considerably different modes of culture aeration and agitation. We attempted to replicate equipment stability, steady state physiology, experimental evolution results, and gene expression patterns obtained with these fermentors.

Since uniformity of dilution rate and aeration are important aspects of the chemostat design, we measured the actual dilution rate across 32 ministats after 15 generations of growth

and found that with a target dilution rate of 0.17 vol/hr (4-5 drops/min) we achieved an average dilution rate of 0.17208 with a standard deviation of 0.0075 across 32 replicates. This range was within our typical tolerance of +/-0.01 vol/hr difference from the target setting, beyond which large-scale changes in gene expression have been observed (Regenberg, Grotkjaer et al. 2006, Castrillo, Zeef et al. 2007). Across 4 ministats the air flow rate was determined to be 307.5 ml/min with a standard deviation of 9.57 ml/min, which suggests that air-flow into the chambers is robust and evenly divided between the 4 chambers and is a value similar to that described for aeration in industrial fermenters (Brauer, Saldanha et al. 2005).

We previously observed recurrent amplification of the high-affinity sulfur-ion transporter *SUL1* in 8/8 sulfate-limited evolution experiments in yeast (Gresham, Desai et al. 2008). Given the consistency of results under this condition we chose sulfate-limitation to test our system's ability. A requisite element of chemostat culture is the need to maintain a constant chemical environment. Fluctuations in the abundance of a limiting nutrient or other changes in the environment typically result in a change in the number of cells in a given culture. We measured optical density as a proxy for cell number (Figure 2A) and measured reproducibility across 16 replicate cultures, finding an average A600 of 1.12 (~109 cells) after ~15 generations of growth with a standard deviation of 0.057 units, or 5.1%. For comparison, measurements taken from 4 replicate cultures grown in the industrial fermentors showed a standard deviation of 2.5%. Cells were well-mixed in the growth chamber: measurements of OD and cell count taken from the effluent track were equivalent to samples taken directly from the culture tube (data not shown). These results demonstrate the robustness of our platform and ability to maintain a constant chemical environment within a similar tolerance as the industrial fermentor.

As a more sensitive readout of physiology, we compared genome-wide steady state gene expression from cultures grown under sulfate limitation in the ministats and in the industrial fermentor. Gene expression in the ministats showed a high degree of similarity across three biological replicates (Figure 2B). We previously noted that for RNA derived from two replicate Sixfors chemostats and co-hybridized to a microarray, expression of 99% of genes fell within a 1.5-fold range, allowing the use of 1.5X as an empirical significance cutoff (Torres, Sokolsky et al. 2007). Gene expression from three replicate ministats, compared pairwise, showed 99% of genes fell within a 1.5-1.7 fold range, comparable to results from the industrial fermentors. The three samples were hybridized to individual arrays and the pairwise ratios calculated afterwards, so these values include inter-array noise in addition to biological noise, potentially overestimating the variation between replicates as compared to the published, co-hybridized results. 138 genes were differentially expressed >1.5-fold in all three ministat replicates as compared to a sample collected from a matched culture grown in the industrial fermentor. Genes decreased in expression in the ministats were heavily enriched for iron metabolism. This signature may reflect the different metal composition of each device configuration: the Sixfors apparatus includes a metal impeller and aeration assembly immersed in the culture, and the media carboys previously used also required metal hardware. The ministat utilizes stainless steel needles, but no other metal components. Genes with increased expression were largely associated with cell membranes, though the biological significance of this association is unclear.

Finally, we tested experimental evolution under these conditions. After 250 generations of sulfate-limited growth, 4/4 clones tested from 4 independent evolving populations showed amplification of *SUL1* as detected by array Comparative Genomic Hybridization (CGH) (**Fig**

2.2C.). This result is consistent with findings in larger volume chemostats over similar time intervals (Gresham, Desai et al. 2008).

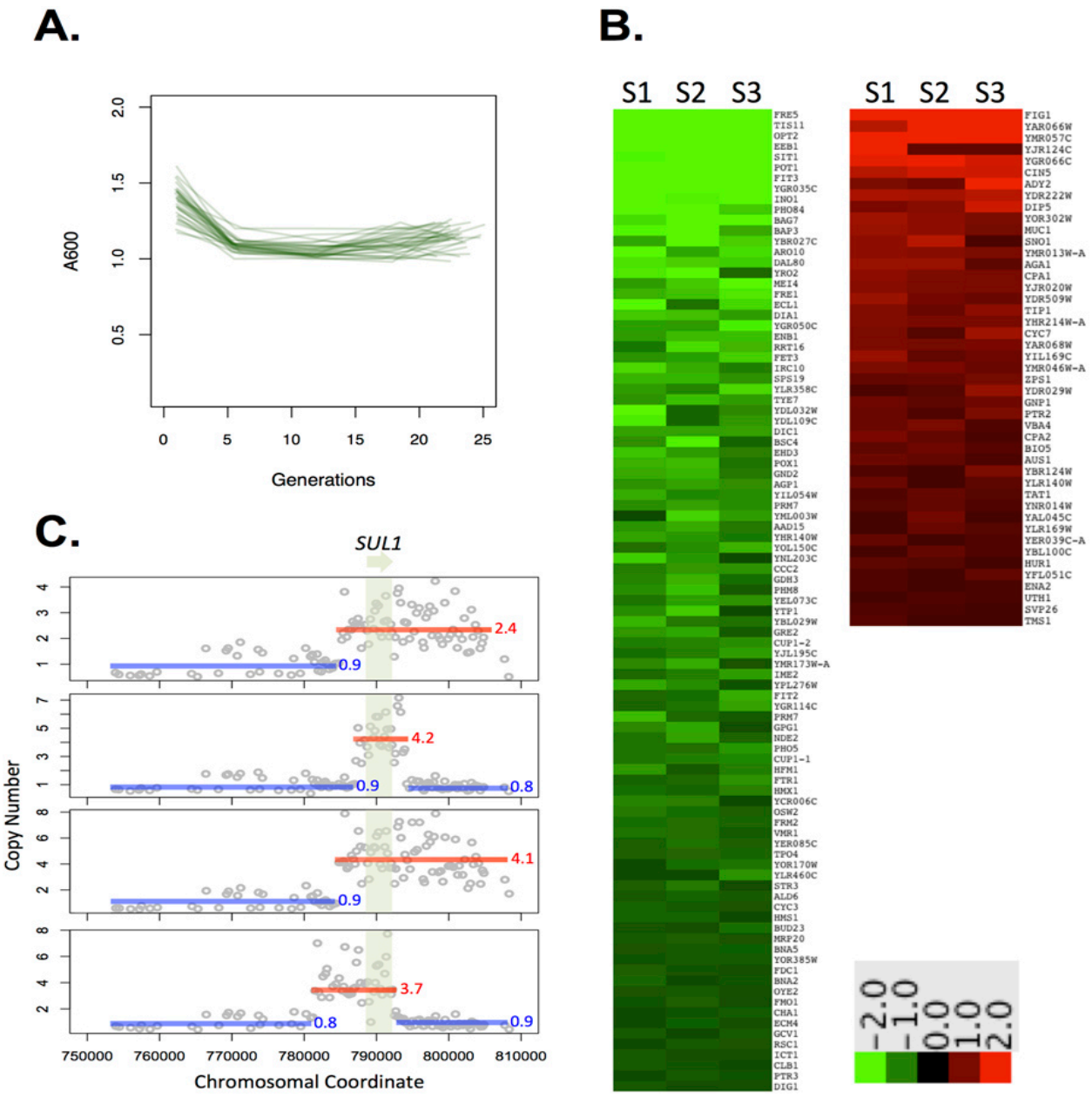


Figure 2.2. A. Experimental data showing that cultures reach equilibrium within ten generations of growth (n = 16). **B.** Expression data for three biological replicates S1-3 sampled during steady state under sulfate limitation compared to a common reference grown in a matched sulfate-limited Sixfors chemostat. **C.** *SUL1* amplifications recovered in ministats after 250 generations of growth in a sulfate-limited environment. Genomic DNA from each evolved clone was compared to ancestral DNA by CGH. Mean copy number was calculated for each amplified region and is shown next to each amplicon. All microarray data are deposited in the GEO database under accession GSE36691.

2.5 DISCUSSION

Chemostat cultivation in the ministats, as with any chemostat, requires attention to detail and trouble-shooting. Since contamination is of great concern in continuous culture experiments, we typically look via microscope for bacterial and fungal contamination upon inoculation and every 50 generations during long-term evolution experiments. To date we have not observed contamination across 96 evolution experiments of greater than 300 generations (data not shown). To test for cross-contamination between ministats and the potential for microbes to colonize the culture chamber by way of the effluent line we ran 16 ministats such that every other ministat was inoculated with yeast as above and the remainder were not inoculated with any culture. The cultures were sampled into a communal waste container, which was emptied every other day. Thus if it were possible for contaminants to enter through the effluent line we likely would have observed that in this experiment. During three weeks and greater than 100 generations of growth in this checkerboard pattern of inoculated and non-inoculated cultures we did not observe growth in non-inoculated ministat culture tubes, suggesting that contamination from outside yeast or other microbes is unlikely to occur in experiments of similar time frames.

Although the ministats were designed to operate in a fashion analogous to commercial chemostats, the modular nature of this arrangement allows for optimization to fit the users' needs

and budget. The peristaltic pump used in this protocol can achieve flow rates between 0.0186 vol/hr to 3.6 vol/hr (data not shown). Increased control of dilution rates could be achieved with alternative pump models. Note that operation at lower dilution rates may require substitution of a higher gauge needle to achieve the same frequency of droplet delivery. Population size is an important consideration for proper design of evolution experiments. The standard dilution rate and nutrient concentration used here provides a relatively large population size (~10^9 cells) of the same order of magnitude as published evolution studies (Kao and Sherlock 2008). Larger or smaller populations could be maintained by changing the working volume or limiting nutrient concentration. Increased mutation supply could also be obtained by working with strains with elevated mutation rates.

The ministats could also be improved over our current design. For instance, condensation can collect on the culture tube walls and can be greatly reduced by using a deep waterbath, incubator, or constant temperature room. Though clumping and wall growth in sulfate-limited cultures appears to be relatively rare, appearing in 5/48 evolution experiments by 300 generations (data not shown), a variety of surfactants are available that may aid in decreasing or delaying this trait. In the event that clumping interferes with adequate culture mixing, increased agitation can be achieved by reducing the number of ways each air pump is divided, or by adding a stirring apparatus. Additional probes for dissolved gas concentration, pH, or other parameters could also be included, as in some other designs (Klein, Schneider et al. 2013).

Despite potential modifications, using the ministats as described in this protocol, we observed highly consistent experimental parameters and reproducible results when comparing our device to reported data for larger volume commercial chemostats. This included reproducibility of cellular physiology as seen through reaching steady-state equilibrium within

10-15 generations (**Fig 2.2A.**) and obtaining similar culture densities at equilibrium. Gene expression patterns were consistent across three biological replicates in ministats and between ministats and commercial large-volume platforms (**Fig 2.2B.**), with the exception of iron metabolism genes. These expression differences are likely caused by changes in metal content of the two devices or improvements in the quality of media ingredients. Our data suggest that ministats will be useful for physiology or competition experiments where a consistent environment is required.

To test if the ministat design is sufficient for experimental evolution applications we evolved cultures under sulfate limitation for 250 generations and used CGH to characterize amplification at the *SUL1* locus - a hallmark of long-term evolution under these conditions in larger volume chemostats (Gresham, Desai et al. 2008). We observed amplification of *SUL1* in clones from 4/4 independent evolution experiments in sulfate-limited media (**Fig 2.2C.**). Taken as a whole these data suggest that ministats are a robust platform that may be useful for a variety of traditional chemostat applications. Though we demonstrated their use in culturing budding yeast, the ministats should also be compatible with other organisms and similar designs have in fact been used for culturing bacteria and other yeast species (Nanchen, Schicker et al. 2006, Ferenci 2008, Klein, Schneider et al. 2013). Furthermore, the smaller culture volume and correlated decreased need for media may make ministats an attractive alternative for experiments requiring expensive or exotic reagents as can be the case in chemical or genetic screens.

2.6 Funding Acknowledgements

Creation of the video was supported by grants from the National Center for Research Resources (5P41RR011823-17) and the National Institute of General Medical Sciences (8 P41 GM103533-17) from the National Institutes of Health. This work was also supported by NSF grant 1120425. MJD is a Rita Allen Foundation Scholar. AWM is supported in part by NIH T32 HG00035.

Chapter 3. THE REPEATABILITY OF EXPERIMENTAL EVOLUTION RESULTS ACROSS MANY REPLICATES

3.1 ACKNOWLEDGEMENTS

I performed the evolution and competition experiments described in this chapter with the help of a series of undergraduates including Bryony Lynch, Annie Young, and Mei Huang. I prepped the sequencing libraries, and performed all analyses including sequencing analysis and gene ontology. I made the figures, interpreted my analysis in terms of known evolutionary trajectories for yeast, and am the author of this chapter. Maitreya and I conceived of the experiments as an interesting application for the chemostat array described in Chapter 2. I had a number of helpful conversations with lab-mates Anna Sunshine and Caiti Smukowski Heil related to sequencing analysis and data interpretation for which I am grateful.

This chapter is part of a manuscript in preparation.

3.1 ABSTRACT

The field of experimental evolution has the central goal of understanding both the genetic cause and phenotypic effect of adaptation in specific environments. Advances in sequencing technologies that enable whole genome characterization of mutations in evolution have led to many useful observations about how specific mutations and interactions between those mutations may shape evolutionary trajectories. However the vast majority of these studies are carried out with limited biological replication, thus hampering accurate assessment of which mutations, genes, or pathways are most important for adaptation in specific environments. Additionally the dynamics of evolution both in terms of rate of adaptation of an evolving population, and the extent to which specific mutations are fixed, are rarely characterized for evolution experiments - thus leaving gaps in what otherwise might be a more holistic understanding of how adaptation occurs. To better understand the effect of adaptation, I performed inline fitness measurements every 50 generations throughout the course of these evolution experiments and observed condition-specific variation in rates at which elevated fitness is achieved. To better understand the genetic cause of adaptation I have performed both array comparative genomic hybridization and whole-genome sequencing on evolved populations after 200 generations of adaptation in glucose-, phosphate-, and sulfate-limited nutrient environments.

During this period of 200 generations, the yeast populations displayed convergent evolution for specific pathways and genes mutated. Most notable is the universal amplification of *SUL1* in 31/31 sulfate-limited populations to a population-average copy number between 2 and 7. Additionally, our whole genome population sequencing identified 769 mutations across 95 populations, many of which had either never been seen before, or were only rarely observed. Roughly 85% of the 769 observed mutations occurred in genes that had not been observed as

mutated in previous studies. Mutated genes include but are not limited to single nucleotide polymorphisms in transporters: *SUL1* in sulfate-limited media and *PHO84* and *PHO89* in phosphate-limited media. Additionally, a few mutations were found with high frequency such as 16 separate predicted loss of function mutations in SAGA complex member *SGF73* in sulfate-limited media; 12 mutations in *GSH1,* which catalyzes the first step of glutathione biosynthesis, and 10 mutations in *BEM2*, which is involved in regulation of bud emergence in glucose limitation. Additionally frequent predicted loss of function mutations in *SIR1, SIR2, SIR3,* and *SIR4* highlight how loss of sirtuin-based sub-telomeric silencing is likely to be beneficial in phosphate limitation. Collectively, these data provide improved quantitation for which specific trajectories may be most important for adaptation in diverse nutrient environments.

3.2 INTRODUCTION

To better characterize the spectrum of mutations responsible for adaptation and their reproducibility across many replicate evolution experiments I employed the use of multiplexed chemostat arrays (Miller, Befort et al. 2013). Chemostats are of great use in experimental evolution because mutations that are observed can be more easily linked to the continuous and invariant selection to which they are exposed through ongoing nutrient limitation (Hoskisson and Hobbs 2005). Glucose, phosphate, and sulfate are three of the four most prevalent components of cells; all are essential for life and commonly used as a limiting nutrient in microbial experimental evolution, particularly in chemostats (Gresham, Desai et al. 2008, Gresham, Usaite et al. 2010, Levy, Kafri et al. 2011). During evolution in chemostats there is a strong selection for mutants with improved capability to transport or utilize a given limiting nutrient, or reduce the function of genes capable of inhibiting cell cycle progression like *RIM15* and *WHI2* that regulate entry

into a quiescent (G_0) state. However, individual studies disagree as to which of these or other mechanisms are most important and have relatively low numbers of biological replicates owing to the difficulty of experimentation and expense of sequencing. Thus, evolutionary mutational screens are far from saturated. With the goal of performing a more comprehensive study of yeast adaptation, in this study I leverage multiplexed arrays of chemostats (see **Chapter 2**) and the decreasing cost of sequencing to interrogate the repeatability of mutations observed after 200 generations of evolution in glucose, phosphate, and sulfate-limited media. Additionally, to better characterize the rate and dynamics of adaptation for these evolving populations I performed regular in-line measurements of fitness. As such, in-depth study of these populations provides a more compressive view into the repeatability and dynamics of adaptation in continuous cultures at an unprecedented scale.

3.3 RESULTS

Evolution in miniature chemostat arrays.

To more comprehensively characterize adaptation across many replicate evolution experiments I used the miniature chemostats described in **Chapter 2**. For this study, I used wild-type FY *MATa* prototrophic yeast. For each culture, a single colony was used to inoculate overnight cultures and ultimately one of 32 chemostats containing glucose-, phosphate-, or sulfate-limited media. These cultures were maintained at a dilution rate of 0.17 ± 0.01 hr^{-1} for 300 generations of continuous growth. Throughout the course of these evolution experiments daily sampling involved measurement of OD600 and creation of glycerol stocks to generate a frozen record of each experiment. Additionally, every 50 generations, coincident with the start of

each competition experiment (described below) I sampled 10 ml of evolved cultures and extracted the genomic DNA for subsequent whole genome sequencing.

In-line measures of fitness reveal condition-specific rate of adaptation and periodic decreases in fitness.

In a chemostat, cells experience a strong selection pressure imposed through limitation of a single nutrient that is essential for growth. In this study I tracked the rate of adaptation for each evolving population measured every 50 generations. To perform these fitness measurements I set up a parallel array of miniature chemostats and inoculated them with a GFP-expressing ancestral strain with neutral fitness. Once the cultures had reached steady-state and stabilized for an additional 5 generations I inoculated 2 ml of sampled evolving culture into the 20 ml GFP+ ancestral culture and tracked the abundance of evolved (GFP-) vs. ancestral (GFP+) strains using cytometry. Competitions generally lasted at least 20 generations after inoculation with cytometry performed every 4 generations (daily). Selection coefficients (s) measured at the initial (generation 15) time-point showed less than a 2% variance across 32 replicates for each condition centered around a selection coefficient of s = 0. (**Fig 3.1.**). This is an encouraging result as it suggests that our measurements of fitness are accurate and precise, since it is unlikely that beneficial mutations present in the inoculum would have had a chance to rise significantly in prevalence; we would thus expect a neutral fitness. It is worthy to note that 15 generations of growth is the first time-point where we confidently assert that cultures have reached steady state based on a plateauing of the culture OD600, and thus we can accurately measure fitness.

By measuring fitness every 50 generations (after the initial generation 15 measurement) for each evolving population, I observed differences between each of the three nutrient

conditions. First, sulfate-limited evolution experiments generally achieved an average increased fitness ($s > 0.2$) in as little as 50 generations of growth in sulfate limitation. Sulfate-limited cultures continued to show the greatest increases in fitness during each 50-generation competition such that they ultimately reached an average selection coefficient of $s = 0.49$ (see **Fig. 3.1A.**). Second, for populations evolved in phosphate limitation I observed an average selection coefficient of $s = 0.12$ after 50 generations. However, these populations continued to rise and nearly reach that of the sulfate-limited populations by generation 300, reaching an average $s = 0.42$. Last, glucose-limited populations achieved only slight increases in fitness across the entirety of the experiment with $s = 0.07$ at 50 generations and $s = 0.11$ after 300 generations of growth (see **Fig 3.1**). Thus we see condition-specific variance in rates of adaptation wherein the strain of haploid laboratory yeast achieved slower and lower increases in fitness under glucose limitation and faster and higher increases in fitness particularly in sulfate limitation but also in phosphate-limited media.

A second main result of studying rates of adaptation for evolving populations is the observation that in more than a third of evolving populations we see sporadic but significant decreases in measured fitness. This phenomenon is observed across all nutrient environments; however, it is most pronounced in sulfate limitation (see **Fig 3.1B,C,D**). One dramatic example of this is 'population S8' grown under sulfate limitation wherein selection coefficients are measured to be $s = 0.41$ at generation 100, $s = 0.07$ at generation 150, and $s = 0.21$ at generation 200. These results are surprising but not entirely unexpected as there is precedence for periodic decreases occurring in chemostat evolution experiments (Paquin and Adams 1983). However, the frequency with which this phenomenon occurs has not been well documented to date. Thus

these data provide a first example that fitness decreases occur quite frequently in chemostat evolution experiments in 3/3 nutrient environments tested.

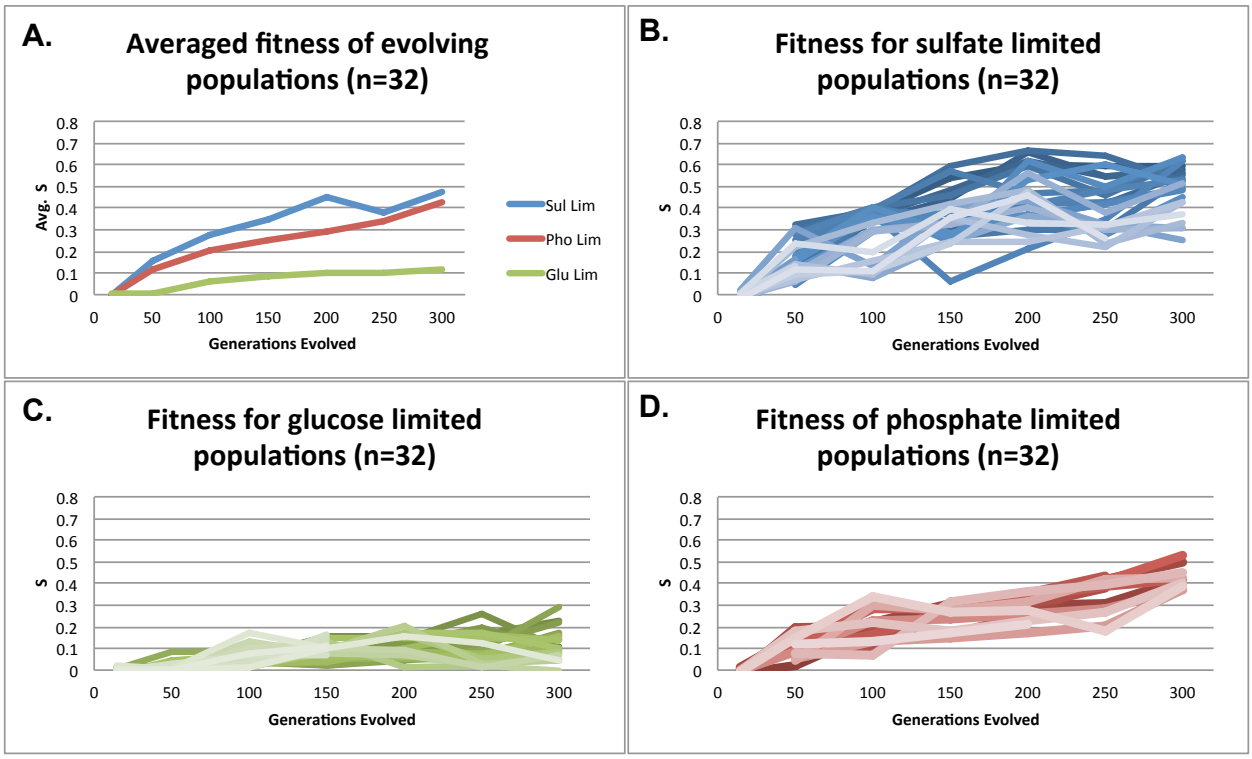


Figure 3.1A,B,C,D. Fitness measurements of adapting populations show condition specific variance in rates of adaptation and periodic decreases in fitness shown in: **A.** Averaged results for each condition **B.** Sulfate limitation. **C.** Glucose limitation. **D.** Phosphate limitation. Although selection coefficients were calculated for populations where flocculation had occurred, they have been removed from this figure.

It is worth noting that strong glass-adherence and cell-clumping phenotypes were observed in 24 evolution experiments, almost entirely between generations 200-300. Selection coefficient measurements for populations where flocculation occurred were removed from figure 3.1 and not included in average selection coefficient measurements in (**Fig 3.1A.**). Additionally, population level sequencing was performed at generation 200 instead of our intended time-point of generation 300 to decrease the likelihood that we would include the causes of flocculation in our analysis of adaptive mutations relevant to the different nutrient environments tested.

Additionally, one population (S32) was contaminated after pump tubing malfunction between generation 150 and 200 and is thus removed from selection coefficient calculations plots after the generation 150 time-point (**Fig 3.1A,B.**).

The spectrum of mutations and mutational classes observed across many replicate chemostat evolutions.

Whole-genome sequencing of evolved populations was performed for all evolution experiments using DNA stocks that were sampled from each chemostat coincident with the generation 200 competition (see the methods section for more information). Since the exact manner of calling and filtering mutations has great impact on the mutations that are called I will briefly explain the general approach I used to call mutations in each population and go into further detail in the methods section. For each population I sequenced the 'generation 200' DNA stock to an average depth of 142.5x using a paired end 300 cycle kit. I then trimmed reads to require a quality score of 30 or greater using Trimomatic (Bolger, Lohse et al. 2014). Next I mapped reads to the reference genome 'SacCer3' using BWA (Li and Durbin 2009). Removal of PCR duplicates was performed using Picard (Li, Handsaker et al. 2009). Mutations were subsequently called using Samtools (Li, Handsaker et al. 2009) and a summary of fold coverage based on mapped reads calculated using the Samtools FLAGSTAT function. The list of mutations called from Samtools was converted into .VCF files using Bcftools, a function contained within the Samtools suite. Mutations with the same coordinates and chromosome number identified in the ancestral FY4 haploid yeast strain were filtered from those identified in each evolved population and annotated using custom Python scripts. The lists of mutations found only in the evolved populations were further filtered in two ways: first, mutations with fewer than 10 alternate reads summed between the forward and reverse direction were filtered out.

Second, given the improbability that identical mutations would occur as *de novo* mutations in replicate populations, these candidates were filtered out until validated via Sanger sequencing. One exception to this filter is that if a given mutation falls into a gene of known benefit in a particular environment then it was not filtered out. A more likely case is that recurrent identical mutations especially in intergenic regions or observed across multiple nutrient environments are representative of either background mutations not called in the sequencing of FY4, or of errors created in the generation of the sequencing library, or sequencing and analysis artifacts such as those seen in repeat rich regions of the genome.

In total, 411 identical mutations were filtered out using these criteria. A high-quality list of 769 mutations remained after filtering for all 95 evolved populations (an average of 8.09 mutations per population). These mutations fell into common mutational classes (**Fig 3.2.**) including truncation mutations such as 'start-lost' and 'early-stop' as well as frame-shifting indels, both of which generally confer a loss of function. Missense mutations were also common and although often detrimental to protein function can also lead to a gain of function or no consequence. Synonymous mutations and 5' upstream promoter mutations were also present although were individually less frequently observed than missense mutations. This class of mutations is less well characterized in the literature although it may result in changes in expression and thus protein function.

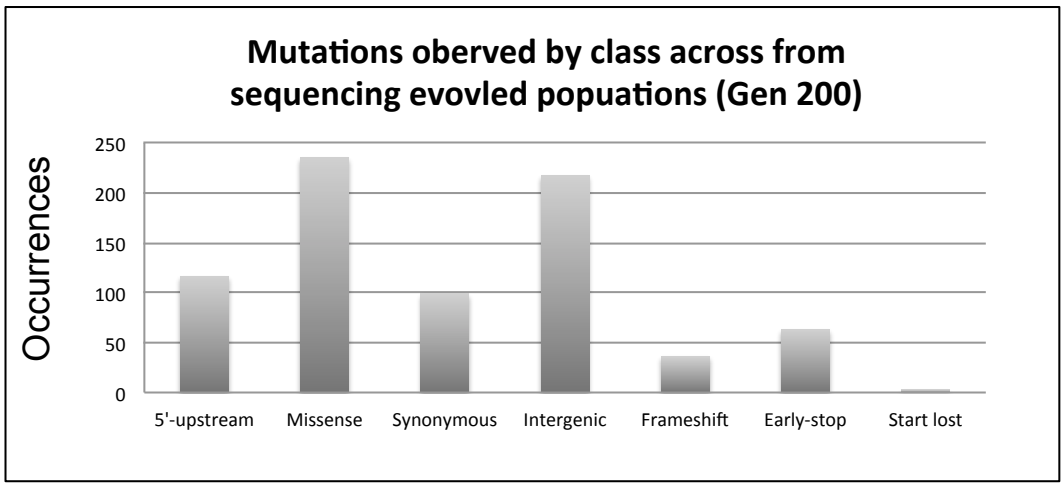


Figure 3.2. Classes of mutations observed amongst 769 mutation and 95 evolution experiments.

Adaptation by nutrient transporter amplification varies greatly based on the selection imposed.

In a chemostat, large populations of cells grow under strong selection for improved growth on a limited nutrient. In both bacteria and yeast grown under nutrient limitations, the target of selection is often a nutrient transport pathway relevant to the imposed selection. For example, mutations in ompF, a gene implicated in membrane permeability, have been observed in *E. coli* grown in lactose-limited conditions (Zhang and Ferenci 1999). In *S. cerevisiae* grown in glucose-, nitrogen-, or sulfate-limited chemostats, amplifications of the glucose, amino acid, and sulfate transporters (*HXT6/HXT7*, *GAP1*, and *SUL1*, respectively) were detected in clones (Brown, Todd et al. 1998, Gresham, Desai et al. 2008, Kao and Sherlock 2008, Levy, Kafri et al. 2011). Measurements from these same studies find that nutrient transporter amplification is generally a large effect mutation that confers 10-50% increased fitness. There are also instances of yeast experimental evolution wherein nutrient transporter amplification is not identified in clones or populations (Gresham, Desai et al. 2008, Kvitek and Sherlock 2013, Payen, Di Rienzi

et al. 2014). Despite these findings on relatively low replicates of the same genotype, it is still unclear how often nutrient transporter amplification occurs across many replicate evolutions.

We observe universal amplification of *SUL1* across 31 evolution experiments at generation 200 (see **Fig 3.3.**). All populations showed an increase in copy number such that the population average was between 2-7. After 200 generations of growth under phosphate limitation we observe 8 cases of *PHO84* amplification from aCGH data, including one case of Chr XIII aneuploidy, the only case of aneuploidy. For populations for which aCGH was not performed I looked at copy number through depth of coverage in sequencing data to get a more comprehensive estimate of the extent of transporter amplification. Copy number variation analysis is still under way for evolved phosphate populations and glucose. For glucose-limited populations, we saw just 2 instances of amplification of the hexose transporter *HXT6/7* in aCGH analysis. Collectively these data provide an example of just how context-dependent nutrient transporter amplification is as a mechanism for adaptive evolution. In sulfate limitation nutrient transporter amplification is likely the most important means through which populations initially adapt. In phosphate limitation, *PHO84* amplification occurred in just 25% of evolutions suggesting other mutations of similar importance may exist in this condition. Data from aCGH suggest that amplification of the hexose transporters *HXT6,7* (which is observed just two times, at a population average copy number of 2) does not appear to be a primary mechanism through which adaptation occurs in these populations. Fitness measurements for *HXT6,7* and *PHO84* amplified strains did not differ significantly from populations that in which these transporters did not amplify, suggesting that for these nutrient conditions large scale rearrangements may be less essential for adaptation and instead evolving yeast may be more reliant on single nucleotide variants and indels.

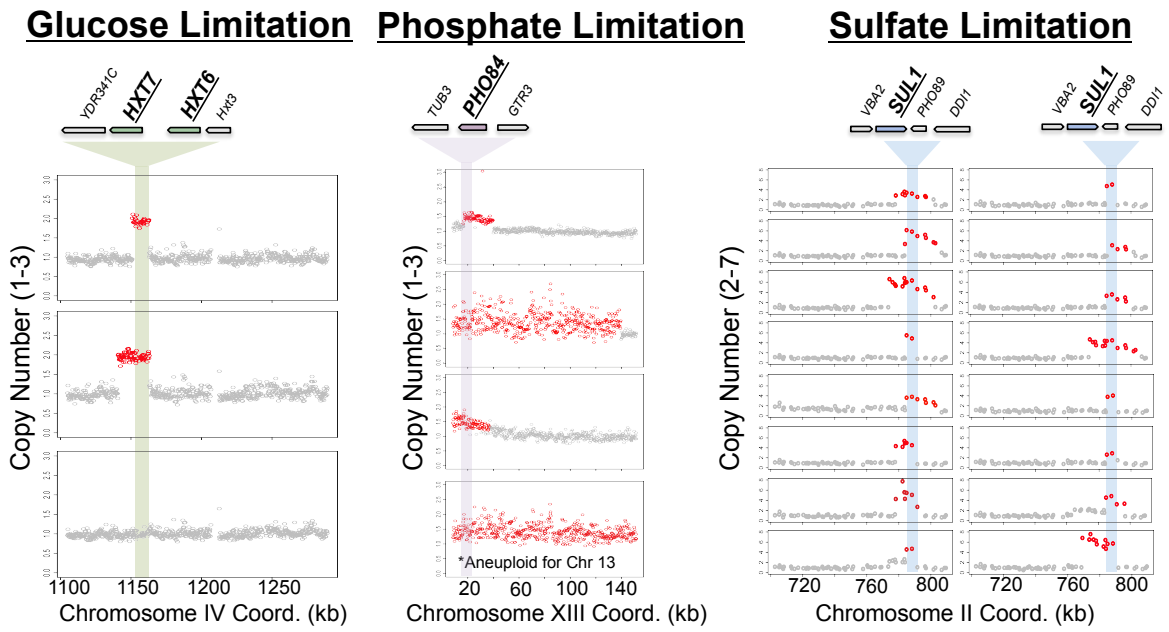


Figure 3.3. Representative results for copy number show huge variation in the extent to which copy number variants are present in the population for generation 200 samples from glucose-, phosphate-, and sulfate-limited environments. Where *SUL1* amplified in 31/31 sulfate-limited evolutions, *PHO84* and *HXT6/7* were only observed to amplify 8 and 2 times, respectively.

Condition-specific trajectories for adaptation across many replicate evolution experiments

We sought to understand the adaptive strategy of yeast growing in three different constant environments for the many replicates tested by defining the main categories of genes that were mutated in each environment. Grouping recurrently mutated genes by pathway or function, we found that in most cases recurrent mutations could be grouped into a few categories based on function or pathway. By performing many repeated evolution experiments in conditions that have already been studied but with fewer replicates, we can identify known categories of genes, pathways, and processes important for adaptation in each nutrient environment as well as hitherto unobserved or unappreciated categories of mutations.

Previous reports have observed that either Hexose transporter amplification *HXT6/7* (Brown, Todd et al. 1998) (which we observe in two evolution experiments) or loss of cell signaling related to nutrient availability and osmoregulation (glucose signaling and transport, cyclic adenosine monophosphate/protein kinase A (cAMP/PKA) and the high osmolarity glycerol (HOG) response pathway) are the primary means through which adaptation occurs for laboratory yeast grown in glucose-limited environments (Kvitek and Sherlock 2013). In this, higher replicate study, we found repeated mutations in the glucose-signaling gene *GPD2* (see **Fig 3.4**). Two separate mutations in *HKR1,* a HOG pathway member, were observed in this study. Although this gene was not mutated in the work of Kvitek *et al.* (Kvitek and Sherlock 2013), it was mutated once in the 40 evolution experiments performed by Greg Lang in batch at roughly 99% prevalence (Lang, Rice et al. 2013), suggesting it was adaptive in rich media. One last category of mutation seen in our study and others are mutations that are likely to be beneficial based on their direct effect on the cell cycle. We saw 6 mutations across two cAMP/PKA signaling genes, *SLN1* and *GPB2.* Loss of function mutations in cAMP/PKA signaling has been observed not just in glucose-limited (Kvitek and Sherlock 2013) but nitrogen-limited environments as well (Hong and Gresham 2014). Both studies posit that loss of function mutations in this pathway would be beneficial because the cAMP/PKA signaling pathway is decreases the rate of cellular proliferation in low-nutrient environments such as that seen in chemostats. Mutations that effect the cell cycle were also seen in the form of repeated *WHI2*, *RIM15*, *NDD1* and *ACE2* mutations (see **Fig 3.4A.**). Collectively, these data show that we find the common classes of mutations reported previously, although in many cases we do not find them to be as prevalent as previous studies might suggest. Our results are surprising in how often alternative means of adaptation occur in our populations.

In addition to known genes or pathways, this study found two genes mutated in roughly 1/3 of the evolved populations in glucose-limited media, although the mechanism by which they influence adaptation is not readily apparent. While mutations in these genes had been seen previously, the fact that they had not been observed repeatedly and have varied functions with no direct link to glucose metabolism or cellular proliferation has prevented substantive discussion of their importance in adaptation. Most notable of these are the 12 missense mutations in *GSH1* (see **Appendix B**), which encodes Gamma glutamylcysteine synthetase, which catalyzes the first step in glutathione (GSH) biosynthesis. This mutation has been seen 4 times in previous studies in diploids (Gresham, Desai et al. 2008, Payen, Di Rienzi et al. 2014, Payen, Sunshine et al. 2015). None of the 12 missense mutations seen in this study match any of the 4 mutations seen in previous studies in diploids; however, it is curious to note the frequent recurrence of missense mutation across the whole length of this protein. Loss of *GSH1* results in a strain that is auxotrophic for glutathione; thus it is unlikely that these mutations would result in a complete loss of function but rather they result in a partial loss or gain of function. Additional studies will be necessary to determine the link between glutathione metabolism and glucose metabolism. A second repeatedly mutated (10 times) gene is *BEM2.* All of these mutations are either early-stop mutations or indels, suggesting that loss of *BEM2* is adaptive in glucose limitation. Although this gene has been seen mutated at least once before (Lang, Rice et al. 2013), it has never been so clearly implicated as a driver of evolution. Indeed, repeated mutation of *GSH1* and *BEM2* suggests that we still have much to learn about adaptation in glucose limitation, arguably the best studied nutrient limitation.

Significantly less is known about adaptation to phosphate limitation. Some previous studies have identified mutation of the high-affinity phosphate transporter *PHO84* as one means

through adaptation occurs under phosphate limitation(Sunshine, Payen et al. 2015). However, due to limited biological replication in any single background, our 32 replicate evolution experiments under phosphate evolution may provide greater clarity toward specific strategies for adaptation in phosphate-limited environments. Based on our whole genome sequencing data of evolved phosphate-limited populations, I observed convergent evolution and frequent mutation of three main categories of genes: phosphate transport, silencing/chromatin remodeling, and control of cell cycle.

In addition to the aforementioned 8 observed *PHO84* gene amplification events (**Fig 3.3**), the most frequently mutated gene was *PHO84*, which acquired 24 identical missense mutations to $PHO84^{L259P}$. The fact that this mutation was found in 75% of our evolution experiments, generally at very high prevalence in the population, is a strong indicator that this SNV is the preferred and primary means to improve phosphate transport in yeast. This specific mutation has been identified before and shown to alter *PHO84* function from genetic linkage mapping for a BY-RM cross. *Perlstein et al*. found that this specific mutation conferred sensitivity to two polychlorinated phenols that uncouple oxidative phosphorylation (Perlstein, Ruderfer et al. 2007). Additionally work in our lab has identified this same mutation twice, once in a diploid FY strains and once in a haploid background FY strain, suggesting that the mutation is beneficial in both ploidy states (Payen, Sunshine et al. 2015). Four phosphate-limited populations acquired *PHO89* mutations, which may act as a secondary means to improve phosphate transport in phosphate-limited media (**Appendix B** and **Fig3.4B.**) Collectively, these data supported the expected mechanism that mutation of genes responsible for transport of phosphate in phosphate-limited media is the primary means by which adaptation to this environment occurs in yeast.

A second expected category seen across multiple nutrient limitations comprises mutations in genes that directly or indirectly influence cell cycle progression. In particular, we obtained loss of function mutations in genes that would otherwise delay cell cycle progression based on environmental stimulus. Six separate mutations were observed in *CKA2,* a Ser/Thr protein kinase with roles in cell growth and proliferation. Mutation of *CKA2* has already been shown to be adaptive in phosphate-limited media through bulk segregant analysis (Gresham, Desai et al. 2008) and has been observed in multiple studies (Gresham, Desai et al. 2008, Payen, Sunshine et al. 2015). Six other mutations were also identified in *RME1,* which is likely to have an indirect effect on cell cycle progression through activation of the G1 cyclin gene *CLN2*. Collectively these results reinforce the idea that cell cycle control is likely an important means of adaptation to growth in chemostats across multiple environments. However, while this category of adaptation persists across nutrient conditions, the specific mutations observed vary between nutrient environments.

One interesting observation from our many replicate evolution experiments in phosphate-limitation is the identification of repeated mutations in genes responsible for silencing and chromatin remodeling. Sirtuin genes *SIR1, SIR2, SIR3,* and *SIR4* are all repeatedly mutated and have an impact on cell cycle progress and silencing of genes at specific locations in the genome, in particular in sub-telomeric regions. As such, it is possible that these genes might act either through their effect on cell cycle control, or their loss might increase expression of genes in sub-telomeric regions, such as the phosphate transporters *PHO84* and *PHO89,* which are located in sub-telomeric regions of chromosomes XIII and II, respectively.

Sulfate limitation saw just two main functional categories of genes that were repeatedly mutated: sulfate import and SAGA/SLIK complex members (**Fig 3.4C.**). In addition to universal

amplification of the *SUL1* locus, we observed 14 missense mutation events to $SUL1^{N250K}$, $SUL1^{N263K}$, and $SUL1^{N263H}$ 8, 4, and 2 times respectively. One population, 'S26', even had $SUL1^{N250K}$ (represented in 2/3 of sequencing reads) and $SUL1^{N263K}$ (represented in 1/3 of sequencing reads), suggesting that these mutations coexisted within separate subpopulations after 20 generations of growth. Observation of SNPs in *SUL1* had not been observed until recently (Payen, Sunshine et al. 2015) and so this stands out as a surprising result made possible by our higher throughput approach to experimental evolution. As with so many other findings in this study, LOF mutations such as early-stop mutations or frameshifting indels in *SGF73* and two other SAGA complex members, *SPT3* and *HFI1,* are scattered throughout previous studies of experimental evolution of yeast in sulfate-limitation. However, observation of 16 mutations in *SGF73* more clearly demonstrates the importance that loss of SAGA function has in adaptation to sulfate limitation (see **Fig 3.4C**). Growth of a *sgf73*Δ strain has been reported to confer a roughly 20% increased fitness, whereas amplification of *SUL1* confers roughly a 40% increase in fitness at high copy number (Payen, Di Rienzi et al. 2014, Payen, Sunshine et al. 2015). Although the SAGA complex has been best characterized for its role in regulating transcriptional activation, SAGA is also required for optimal transcription elongation, mRNA export, and perhaps nucleotide excision repair (Baker and Grant 2007). Further study is required to determine the exact mechanism through which loss of SAGA function is beneficial in sulfate limitation. Collectively, these data are an example of how performing many evolution experiments more conclusively identifies the categories of mutations that are most important for adaptation in a given environment based on repeated mutations within specific genes and pathways.

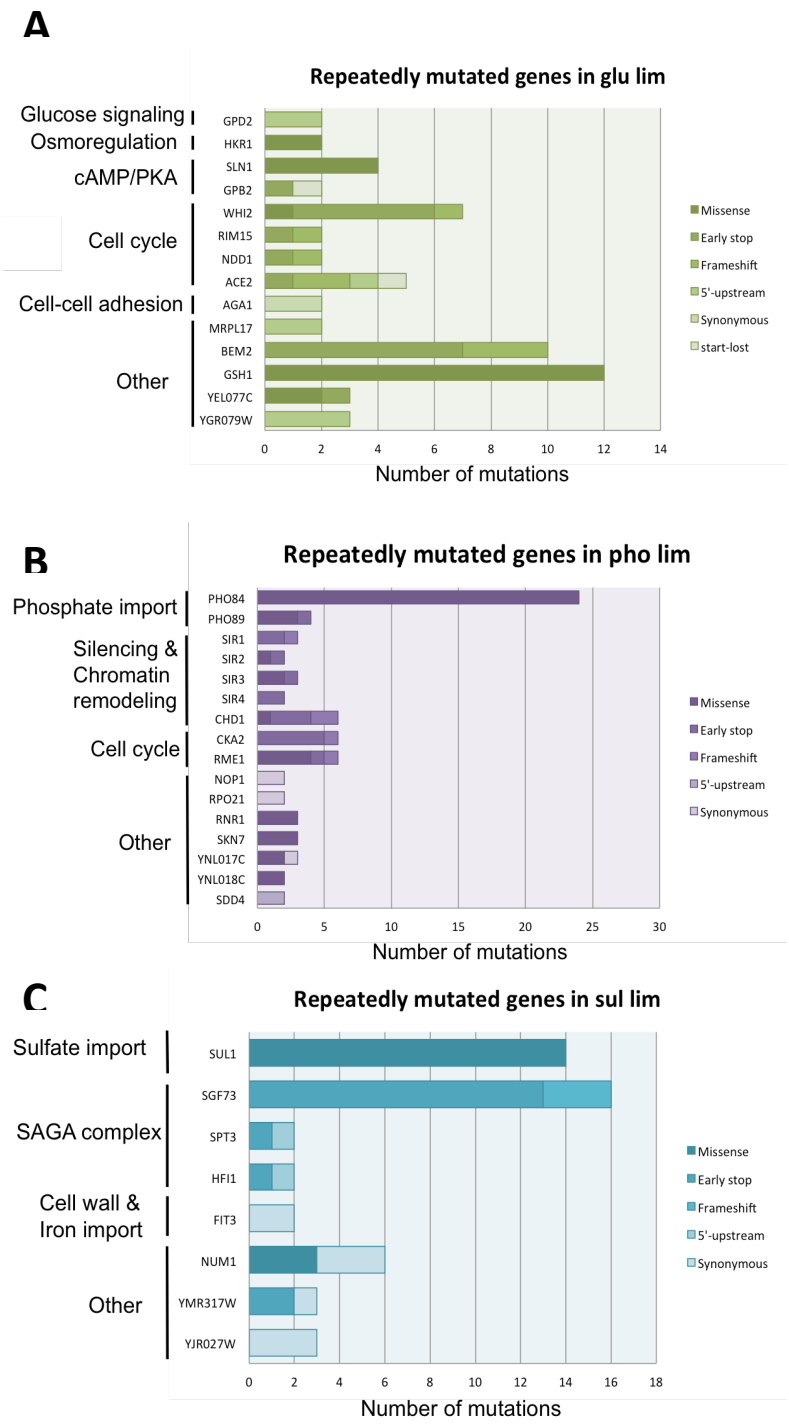


Figure 3.4. Identity and mutational class of recurrently mutated genes across all 95 evolution experiments. Only genes with two or more identified mutations are included; bars are colored according to the mutational class to which each mutation belongs. **A.** Glucose limitation. **B.** Phosphate limitation. **C.** Sulfate limitation.

3.4 DISCUSSION

The repeatability of evolutionary trajectories is one of the most important observations in experimental evolution research. Thousands of genes are available for mutation and mutations accumulate at random and subsequently rise in frequency from a single mutational event when they are beneficial. Repeatedly observing mutations in a gene or pathway is one of the most common ways to distill meaning from the seemingly random set of mutations that arise in microbial evolution experiments. Repeatability can even be used to prioritize mutations observed in evolution experiments. Identical mutations seen in replicate evolutions are suspect and must be validated, given the improbability that identical mutations would arise in a 12 megabase genome at high prevalence in two or more separate evolving populations. As such, identical mutations are filtered from mutations called outright, with the rationalization that they are more likely background or easily generated PCR or sequencing artifacts (such as is the case with repeat rich regions) rather than *de novo* adaptive mutations. Conversely, the requirement that a given mutation is seen repeatedly at two or more time-points in an evolution experiment has been used to suggest that the mutation is legitimate and others that do not repeat at two or more time-points are filtered out (Lang, Rice et al. 2013). Lastly and most importantly, co-occurrence of different mutations within the same gene or functional pathway is thought to be a strong indicator that the mutations are beneficial. It is for this reason that my analysis has focused on repeatability.

Two findings in this study seem inconsistent with previous studies. One is that we found 10 deleterious mutations in the gene *BEM2*, a Rho GTPase activating protein involved in the control of cytoskeleton organization and cellular morphogenesis. Given the frequency at which glucose-limited evolution experiments have been performed in the past and subjected to whole

genome sequencing, I would expect this mutation to have been found in roughly 1/3 of the experiments performed by others using similar strains and growth conditions, but this mutation is not observed elsewhere. A second place where our results diverge from previous studies is just one glucose-limited population, 'G4', had a loss of function mutation in $mth1^{E303*}$ represented above our filtering cutoffs in this study. Previous work suggests that we might have expected to see many mutations in this gene. One plausible explanation for the low frequency of *mth1* mutations is negative epistasis or sign epistasis between the mutations that are observed in our study in the commonly mutated genes *BEM2* and *GSH1*. In this study and that of others, these mutations do not appear to co-occur, so this observation may not be without merit.

Despite the importance of repeatability of evolutionary trajectories for understanding which mutations are most important for adaptation, most continuous culture-based evolution experiments are performed with few replicates per genotype tested. It is often the case that previous studies found mutations once or twice that in our hands were observed in 16/31 populations, such as loss of function mutations in *SGF73*. This same finding is also true for Sirtuin genes in phosphate limitation, which were found in clones from one population from a previous study published by our group (Gresham, Desai et al. 2008). Each of the Sirtuin genes *SIR1, SIR2, SIR3*, and *SIR4* were mutated in this study only 2 or 3 times, which implies that one must indeed perform dozens of replicate evolution experiments to reliably observe repetition for mutations for these genes and pathways. However, loss of silencing associated with Sirtuin function appears to be one of the main categories of mutations in phosphate limitation. Collectively, these findings suggest that greater replication is needed to predict what mutations in genes and pathways are most important for adaptation.

3.5 METHODS

Growth in chemostats.

Wild-type FY *MATa* prototrophic yeast was used for all experiments in this study unless otherwise stated. For each culture, a single colony was inoculated into glucose-, phosphate-, or sulfate-limited media and grown as overnight cultures. 100 μL was inoculated into miniature chemostats (Chapter 2) containing 20 mL of the same medium. 30 hours post-inoculation, the peristaltic pumps that controlled flow of media to the chambers were turned on and used to maintain a dilution rate of 0.17 ± 0.01 hr^{-1} for 300 generations of continuous growth.

Throughout the course of these evolution experiments, daily sampling involved creation of glycerol stocks to generate a frozen record of each experiment. Additionally, every 50 generations, coincident with the start of each competition experiment, I sampled 10 ml of evolved cultures and extracted the genomic DNA for subsequent whole genome sequencing.

Determination of selection coefficients for evolving cultures.

Competitions were carried out against a GFP-expressing haploid FY ancestral strain. To assay the fitness of evolving populations a separate chemostat was inoculated with this GFP+ reference strain. Once the GFP+ ancestral strain reached steady-state (after 15 generations of growth) 2 ml of the evolving (GFP-) culture was inoculated into the GFP+ culture at steady-state under the same conditions. Competitions generally ran for 20 or more generations, with daily sampling (every ~ 4 generations). Quantification of how prevalent GFP- (evolved) and GFP+ (ancestral) cells were at each time-point was measured using a BD Accuri C6 flow cytometer. The data were plotted and selection coefficients calculated from the slope of the ln(GFP-/GFP+)/generations elapsed.

Sequencing methods

DNA was extracted using the Yeastar Genomic DNA extraction kit. Genomic DNA libraries were prepared for sequencing using the Nextera sample preparation kit (Illumina). Sequencing libraries were quantified on a Qubit Fluorometer (Invitrogen) and submitted for 150 bp paired end sequencing on an Illumina NextSeq. For each population I sequenced the 'generation 200' DNA stock to an average depth of 142.5x using a paired end 300 cycle kit. I then trimmed reads to require a quality score of 30 or greater using Trimomatic (Bolger, Lohse et al. 2014). Next I mapped reads to the reference genome 'SacCer3' using BWA (Li and Durbin 2009). Removal of PCR duplicates was performed using Picard (Li, Handsaker et al. 2009). Mutations were subsequently called using Samtools (Li, Handsaker et al. 2009) and a summary of fold coverage based on mapped reads calculated using the Samtools FLAGSTAT function. The list of mutations called from Samtools was converted into .VCF files using Bcftools, a function contained within the Samtools suite. Mutations with the same coordinates and chromosome number identified in the ancestral FY4 haploid yeast strain were filtered from those identified in each evolved population and annotated using custom Python scripts. The lists of mutations found only in the evolved populations were further filtered in two ways: firstly mutations with fewer than 10 alternate reads summed between the forward and reverse direction were filtered out. Additionally given the improbability that identical mutations would occur as *de novo* mutations in replicate populations, these candidates were filtered out until validated via Sanger sequencing. One exception to this filter is that if a given mutation falls into a gene of known benefit in a particular environment then it was not filtered out.

3.6 FUNDING ACKNOWLEDGEMENTS

AWM was funded by both the (GTG) genome training grant and IDTG interdisciplinary training grant in cancer co-administered by UW and the Fred Hutchinson Cancer Research Center.

Chapter 4. NOVELTY BY NECESSITY: LOSS OF SULFATE TRANSPORT REPEATEDLY SELECTS FOR MUTATIONS IN AN UNCHARATERIZED TRANSPORTER *YIL166C*

4.1 ACKNOWLEDGEMENTS

Ivan Liachko made the *sul1*Δ*, sul2*Δ and *sul1*Δ *sul2*Δ strains described in this chapter. I created the *yil166c*Δ*,* and *yil166c*Δ *sul1*Δ *sul2*Δ strains. Undergrads Erica Alcantara, Thao Nyguen, and I performed the *sul1*Δ evolution experiments and associated competitions. 16 of the sulfate-limited evolution experiments described in (**Chapter 3**) were also characterized via aCGH for comparison to the *sul1*Δ evolution experiments. I performed the batch evolution experiments. I prepared the sequencing libraries and performed all sequencing analyses including SNP/INDEL calling and CNV analysis.

I performed a series of ^{35}S uptake assays that showed that *YIL166C* encoded a low-affinity sulfate transporter. However, restrictions on how much radioactive sulfate can be disposed of per quarter hindered these experiments and eventually led us to seek out collaborators to further characterize *YIL166C*-relevant genotypes and associated evolved clones. A natural collaborator presented itself in the form of Sylvester Holt from Johan Thevelein's lab at KU Leuven in Belgium. Sylvester had also noticed that *YIL166C* encoded a low-affinity sulfate transporter using a candidate-based approach to identify sulfate transporters based on the observation that *YIL166C* had increased expression in sulfate-limited chemostats. I generated the *yil166c*Δ *sul1*Δ *sul2*Δ *his3* strain used for the sulfate assays and cloned the evolved *YIL166C* alleles into pRS413. Interpretation of our findings and the writing of both this chapter and a manuscript based on these data were prepared by me. Erica Alcantara and I worked together to calculate doubling times from the Omnilog screen described in this chapter, and performed subsequent experiments to test transport of both nitrogen and sulfur sources.

This chapter is part of a manuscript in preparation.

4.2 ABSTRACT

Despite novelty's importance in evolution, it is rarely observed in microbial or metazoan experimental evolution. In this study, aCGH and whole genome sequencing were used to investigate the evolution of a novel sulfate transporter in *Saccharomyces cerevisiae* that can compensate for the loss of yeast's canonical sulfate transporter genes *SUL1* and *SUL2*. To explore how proteins with novel function might arise in yeast, I performed three independent evolution experiments using sulfate transport-deficient (*sul1*Δ *sul2*Δ) strains in decreasing concentrations of sulfate-limited media in serial batch cultures. All three independent evolution experiments yielded sulfate transport-proficient populations capable of growth on sulfate-limited media (3mg/L ammonium sulfate). Clones derived from these populations were subjected to whole genome sequencing and found to contain 3 different missense mutations in the 2nd MFS domain of a previously uncharacterized transporter gene, *YIL166C*. Introduction of these *YIL166C* alleles in a *sul1*Δ *sul2*Δ *yil166c*Δ strain was sufficient for growth on sulfate minimal media whereas the wt allele was not, demonstrating that these alleles had become efficient sulfate transporters. Phenotyping of *yil166c*Δ and wt strains suggest that *YIL166C* encodes a broad-specificity transporter that mediates transport of a variety of sulfate and nitrogen sources. Subsequent mechanistic studies showed that the ancestral Yil166c has a high affinity for sulfonates (K_m of 109μM) and vastly lower affinity for sulfate. The evolved alleles of $YIL166C^{Q338K}$,$YIL166C^{Y428H}$, and $YIL166C^{Y454H}$ all resulted in pleiotropic improvements in transport of sulfate as well as isethionate. Thus, we see convergent compensatory evolution in which an apparently broad specificity transporter may be repurposed to compensate for the loss of sulfate transport proficiency. Collectively, this work is an example of how we may be able to

use suppressor screens carried out across evolutionary time scales to reveal proteins capable of acquiring new function and thus better understand the potential for novelty in yeast evolution.

4.3 INTRODUCTION

Adaptive evolution proceeds by accumulation of beneficial mutations for the environment in which cells are exposed. However, most mutations are either neutral or deleterious. These mutations can still appear at high frequency in a population via hitchhiking or drift. Deleterious mutations generate a selective pressure for improved fitness through compensatory mutation, which may be a mechanism for population divergence. Recent studies contain conflicting results as to whether the genetic basis of compensatory evolution is divergent or convergent in nature. For example, Szamecz *et al.* (Szamecz, Boross et al. 2014) found that across 180 deletion mutants with poor growth, the majority (68%) rapidly acquired mutations that compensated for the original gene deletion. Of these, the vast majority of compensatory mutations were determined to be in non-paralogous and functionally non-similar genes. These results are in contrast to cases of convergent evolution where mutation of a paralog has been shown to compensate for the loss of a specific genes (Hughes, Roberts et al. 2000). These latter findings suggest that compensatory evolution occurs via mutations amongst paralogs or genes with function similar to the one removed (de Kok, Nijkamp et al. 2012, Tenaillon, Rodriguez-Verdugo et al. 2012, Laan, Koschwanez et al. 2015). It is unclear whether certain classes of deleterious mutations are more likely to promote convergent or divergent compensatory outcomes.

Long-term growth across a range of wt laboratory yeasts repeatedly selects for amplification of the high-affinity sulfate transporter *SUL1,* detected in as few as 50 generations

of growth ((Gresham, Desai et al. 2008), (Payen, Di Rienzi et al. 2014)) and amplification of *SUL1* is virtually universal across all evolution experiments performed to date (see **Chapters 2 and 3**). Sulfate transport in yeast is believed to be mediated by only two high-affinity sulfate transporters: *SUL1* and its paralog *SUL2* (Cherest, Davidian et al. 1997). Loss of a single sulfate transporter does not prevent growth at low sulfate concentrations. Loss of both transporters prevents growth unless the growth media is supplemented with high concentrations of ammonium sulfate. Given the universality of *SUL1* amplification in sulfate-limited media, I chose to study compensatory outcomes in a genetic background lacking one (*sul1*Δ) or both (*sul1*Δ *sul2*Δ) sulfate transporters evolved in sulfate-limited environments.

4.4 RESULTS

Compensatory evolution and unanimous convergent evolution for partial loss of sulfate transport amplification.

In this study we have used methods of experimental evolution to interrogate sulfate transport and compensation strategies for loss of sulfate transport in *Saccharomyces cerevisiae*. Previous reports have implicated sulfate transporter amplification as a primary means of adaptation to extended growth in sulfate-limited media (Gresham, Desai et al. 2008, Payen, Di Rienzi et al. 2014). However, to date only a few biological replicates for a given genetic background have been tested. To get a baseline for how repeatedly *SUL1* amplifies in sulfate-limited media we evolved 16 separate populations of haploid wt *MATa* (FY) yeast in an array of multiplexed chemostats for 200 generations (see **Chapter 3**). aCGH data from all 16 populations showed amplifications that contain *SUL1* (**Fig 4.1A.**). Evolution of *sul1*Δ strains for 200 generations selects for amplification of *SUL1's* paralog *SUL2* in 4/4 cases tested (**Fig 4.1B.**). The

pervasiveness of *SUL2* amplification in the *sul1*Δ background highlights the importance of how pre-existing genetic redundancy can have a major impact on compensatory evolution.

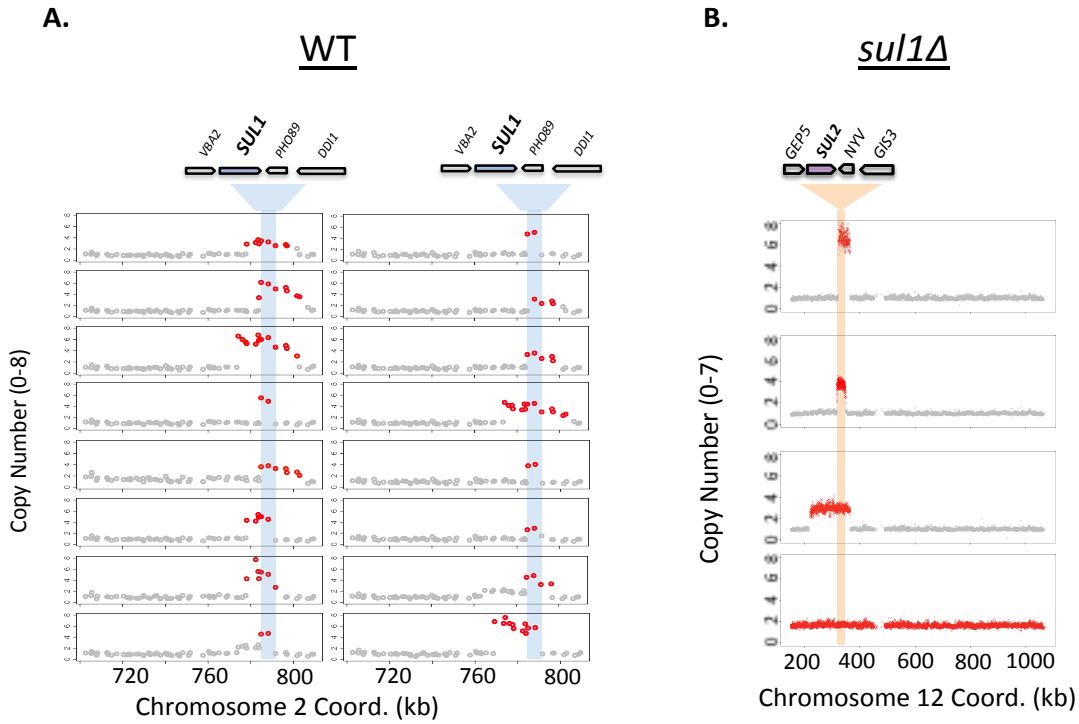


Figure 4.1. aCGH analysis of **A.** wt yeast and **B.** the *sul1*Δ strain background highlight universal amplification of sulfate transporters evolved in sulfate-limited media.

Laboratory evolution of *SUL1 SUL2* independent growth on sulfate-limited media.

To better understand compensatory mutations that would allow for loss of sulfate transport a *sul1*Δ *sul2*Δ knockout strain was constructed and evolved on incrementally decreasing concentrations of ammonium sulfate. These experiments are distinct from a classic suppressor screen in that they are carried out over many generations and allow for multiple mutations to arise. To date, *SUL1* and *SUL2* are the only two known transporters of sulfate in *S. cerevisiae* (Cherest, Davidian et al. 1997). Consistent with earlier reports, the *sul1*Δ *sul2*Δ strain

did not show growth on media containing low concentrations of sulfate after two weeks, whereas the reference wt strain FY4 grew normally. A range of ammonium sulfate concentrations was tested and 300 mg/L (~2mM) ammonium sulfate resulted in growth of the *sul1*Δ *sul2*Δ strain within a week, while lower concentrations resulted in no growth within two weeks. This concentration is high enough that sulfate might not be limiting to a transport-proficient strain thus complicating the use of chemostats for this purpose. Thus, to select for growth on lower sulfate concentrations small aliquots of the batch cultures were first grown on minimal media containing 300 mg/L ammonium sulfate. As soon as visible growth was evident, a 100 µl aliquot of each culture was passaged into a number of culture tubes containing 10 ml of fresh minimal medium supplemented with a range of ammonium sulfate concentrations (300 mg/L, 200 mg/L, 100 mg/L, 50 mg/L, 6 mg/L, and 3 mg/L). As soon as visible growth was observed (usually within 3-4 days) cultures were passaged into equivalent and reduced sulfate concentrations, each time carrying forward only the culture that grew at the lowest sulfate concentration tested (**see 4.2A.**). Hemocytometer-based measurements of cell density were performed for all passaged cultures and used to calculate the number of generations transpired. Crucial to the experimental design is that for each culture that was transferred, diagnostic PCR was performed to make sure that the evolving population was *sul1*Δ *sul2*Δ and not wt contaminant. After ~170 generations (roughly two months of growth), three populations were found to be capable of visible growth in 3 mg/L sulfate-limited media. Frozen glycerol stocks were generated from each sulfate-proficient population. Both populations and clones were frozen as glycerol stocks and utilized in subsequent experiments.

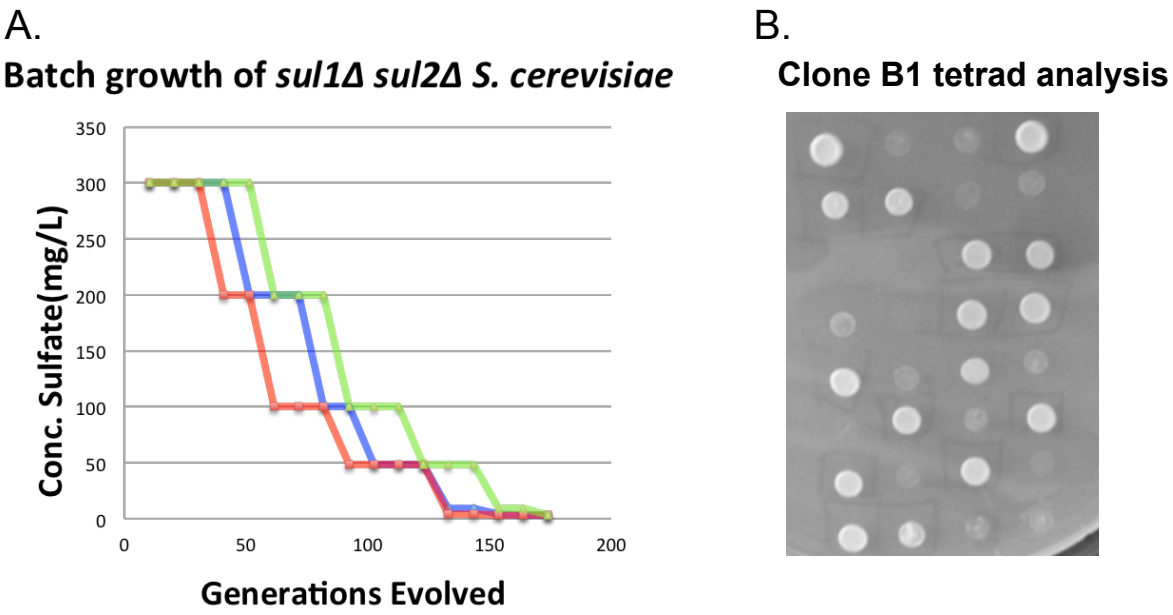


Figure 4.2A,B. **A.** Growth of *sul1*Δ *sul2*Δ strains in sulfate-limited media results in sulfate transport-proficient populations by ~170 generations. **B.** Tetrad analysis of clone 1 from population B shows 2:2 segregation, suggesting a single locus confers ability to grow on sulfate-limited media (3 mg/L ammonium sulfate). Subsequent Sanger sequencing of sulfate transport-proficient segregants shows that they all have the *YIL166C*^{Q338K} allele described below.

Whole genome sequencing of sulfate-transport proficient *sul1*Δ *sul2*Δ strains identify mutations in an uncharacterized transporter *YIL166C*

Sulfate transport-proficient clones from evolved *sul1*Δ *sul2*Δ populations were subjected to whole genome sequencing to identify causal mutations. Clones were sequenced to an average depth of coverage of between 80-160 fold (**Fig 4.3A.**). The sequencing reads were mapped to the S288C reference and after subtracting mutations specific to the *sul1*Δ *sul2*Δ ancestral strain, a total of 12 unique mutations were identified across all clones (**Fig 4.3B.**). A summary of SNVs and INDELS identified for each clone is listed in **Table 4.1**. Clones from all three cultures

contain missense mutations, either Q338K, Y428H, or Y454H in a previously uncharacterized non-essential transporter, *YIL166C*.

Table 4.1. Mutations Identified in evolved sulfate transport proficient clones.

Clone	Chrom	Coord	Ref	Alt	Mutation Type	Sys Name	aa change	Common name
A1	chrII	389421	GCTCTCT	GCTCT	ARS	ARS213		
A1	chrV	499602	G	C	coding-nonsynonymous	YER161C	Y249*	SPT2
A1	chrIX	31285	A	G	coding-nonsynonymous	YIL166C	Y428H	
A1	chrIX	38113	G	T	coding-nonsynonymous	YIL162W	M243I	SUC2
A2	chrV	499602	G	C	coding-nonsynonymous	YER161C	Y249*	SPT2
A2	chrIX	31285	A	G	coding-nonsynonymous	YIL166C	Y428H	
A2	chrIX	38113	G	T	coding-nonsynonymous	YIL162W	M243I	SUC2
B1	chrII	517400	C	A	coding-nonsynonymous	YBR140C	E3077*	IRA1
B1	chrV	509042	T	C	coding-synonymous	YER164W	N1217N	CHD1
B1	chrV	499696	TC	T	coding-nonsynonymous	YER161C	E218indel	SPT2
B1	chrV	509044	C	T	coding-nonsynonymous	YER164W	P1218L	CHD1
B1	chrIX	31555	G	T	coding-nonsynonymous	YIL166C	Q338K	
B1	chrX	415757	G	T	intergenic	NA	NA	
B2	chrII	517400	C	A	coding-nonsynonymous	YBR140C	E3077*	IRA1
B2	chrV	499696	TC	T	coding-nonsynonymous	YER161C	E218indel	SPT2
B2	chrV	509042	T	C	coding-synonymous	YER164W	N1217N	CHD1
B2	chrV	509044	C	T	coding-nonsynonymous	YER164W	P1218L	CHD1
B2	chrIX	31555	G	T	coding-nonsynonymous	YIL166C	Q338K	
B3	chrII	517400	C	A	coding-nonsynonymous	YBR140C	E3077*	IRA1
B3	chrV	499696	TC	T	coding-nonsynonymous	YER161C	E218indel	SPT2
B3	chrV	509042	T	C	coding-synonymous	YER164W	N1217N	CHD1
B3	chrV	509044	C	T	coding-nonsynonymous	YER164W	P1218L	CHD1
B3	chrIX	31555	G	T	coding-nonsynonymous	YIL166C	Q338K	
B3	chrXII	605398	T	C	tRNA	tI(UAU)L	NA	
C1	chrIX	31207	A	G	coding-nonsynonymous	YIL166C	Y454H	
C2	chrIX	31207	A	G	coding-nonsynonymous	YIL166C	Y454H	
C3	chrIX	31207	A	G	coding-nonsynonymous	YIL166C	Y454H	

These *YIL166C*^{Q338K}, *YIL166C*^{Y428H}, and *YIL166C*^{Y454H} mutations are predicted to fall into the transmembrane region of the second MFS domain of *YIL166C* (**See Fig 4B**). We hypothesize that

these missense mutations confer an increased ability to transport sulfate in these evolved clones. Regulation of transcription may be a second important target for adaptation. In 2/3 of the evolved *sul1*Δ *sul2*Δ populations I observed mutations in the negative transcriptional regulator *SPT2.* These include an early stop allele $SPT2^{Y249*}$ and a frame-shifting INDEL at E218. Both likely result in loss of function mutations for the negative transcriptional regulator *SPT2* in populations 'A' and 'B', respectively. In these evolving *sul1*Δ *sul2*Δ populations I observed mutations commonly associated with adaptation to other nutrient environments. For instance, *CHD1* was seen as repeatedly mutated in phosphate limitation (**Chapter 3**) and mutations in *IRA1* have been shown to be adaptive in glucose limitation (Kao and Sherlock 2008). Collectively, these data implicate both global transcriptional regulation and RAS/PKA signaling as a potential alternative means through which *sul1*Δ *sul2*Δ strains may adapt to sulfate-limited environments.

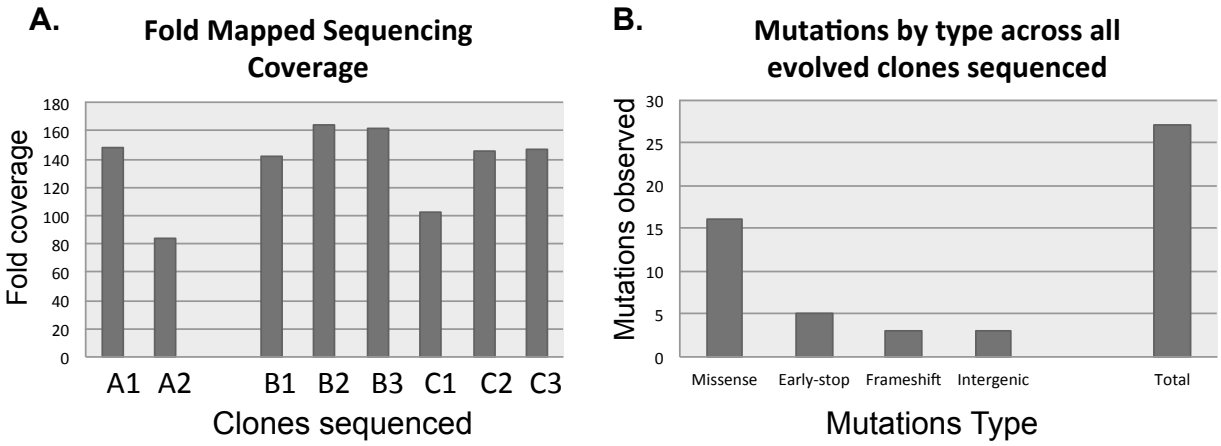
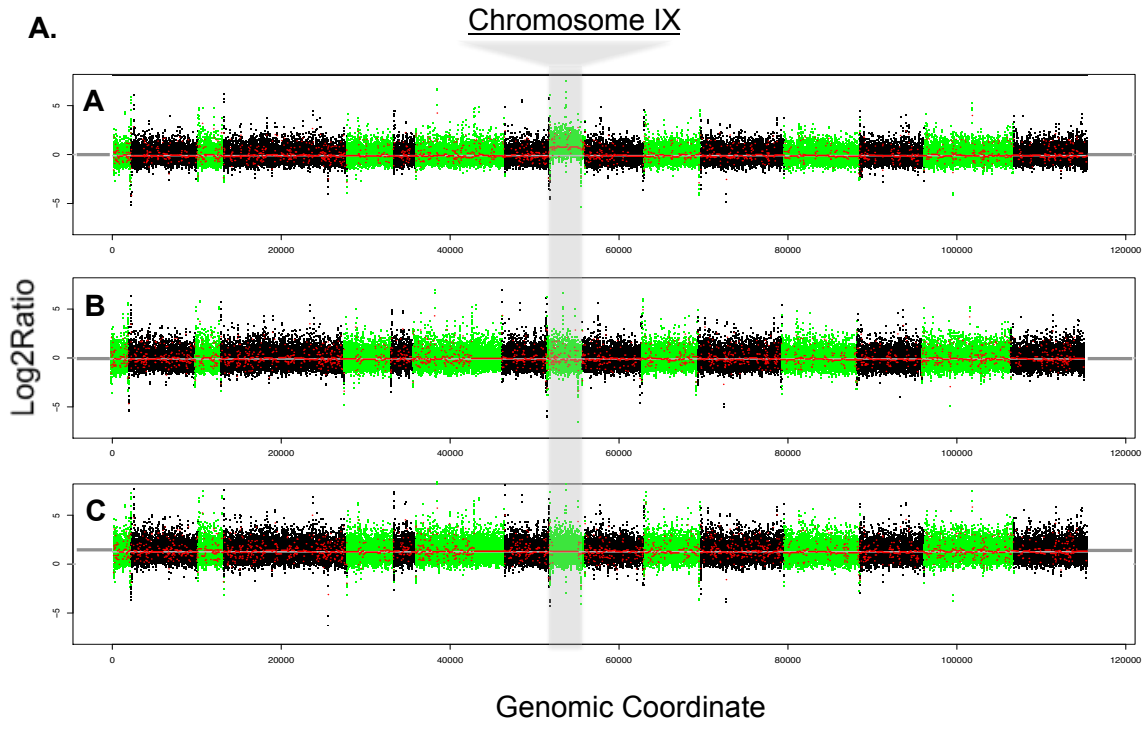


Figure 4.3A,B. A. We achieved between (80-160)-fold coverage for each sequenced clone with the exception of clone 'A3' which is being re-sequenced. **B.** Coding non-synonymous mutations account for nearly all mutations seen in sulfate transport proficient evolved *sul1*Δ *sul2*Δ clones.



B.

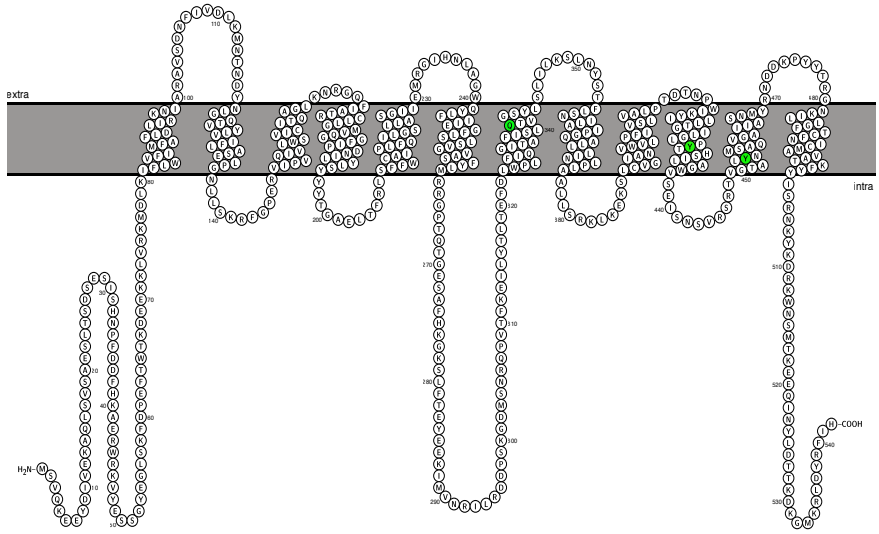


Figure 4.4A,B. A. Extended growth of a *sul1*Δ *sul2*Δ strain repeatedly selects for mutations in novel sulfate transporter *YIL166C*. **A.** Whole genome sequencing reveals copy number variation and missense mutations in *YIL166C* for sulfate proficient clones visualized using DNAcopy. **B.** Protter plot (Omasits, Ahrens et al. 2014) of *YIL166C* with mutated amino acids highlighted in green shows that missense mutations fall into transmembrane regions.

Copy number variation appears to be a second important means for adaptation. A comparison of depth of coverage of DNA sequencing paired with Sytox staining (**Fig 4.4A.**) shows an extra copy of chromosome IX for clones from population 'A' and a complete diploidization event for population C. As *YIL166C* is present on chromosome IX the increased copy number for this whole chromosome in population 'A' may further implicate mutation of *YIL166C* as an important compensatory mutation for this *sul1*Δ *sul2*Δ strain. For both populations 'A' and 'C' mutations in *YIL166C* preceded the copy number variation as evidenced by the fact that the alternate allele is represented in 100% of sequencing reads. While clones from populations 'A' and 'C' could not be used for tetrad analysis and subsequent phenotyping, population 'B' was amenable to such approaches and we observe that ability to grow on sulfate-limited media segregated 2:2 when the strain was back-crossed to a *sul1*Δ *sul2*Δ *MATα* strain (see **Fig 4.2B.**). This finding suggests a single locus is most important for growth in sulfate-limited media. Subsequent Sanger sequencing of the transport-proficient clones showed that they all contained the $YIL166C^{Q338K}$ mutation. Given each population's periodic and incremental improvement for growth at lower sulfate concentrations it is likely that other mutations may also be adaptive although of lower importance.

Expression of evolved alleles of *YIL166C* alone is sufficient for efficient sulfate transport.

To verify whether mutation of *YIL166C* alone is sufficient to allow for growth on sulfate-limited media, I expressed the mutated alleles in a non-evolved background (*sul1*Δ *sul2*Δ *yil166c*Δ *his3*) using the CEN expression vector pRS413. Indeed expression of the evolved *YIL166C* alleles increased growth in sulfate-limited media (**Fig 4.5.**) The implication is that these

missense mutations of *YIL166C* convert it from an either inefficient or non-transporter of sulfate into an efficient transporter of sulfate.

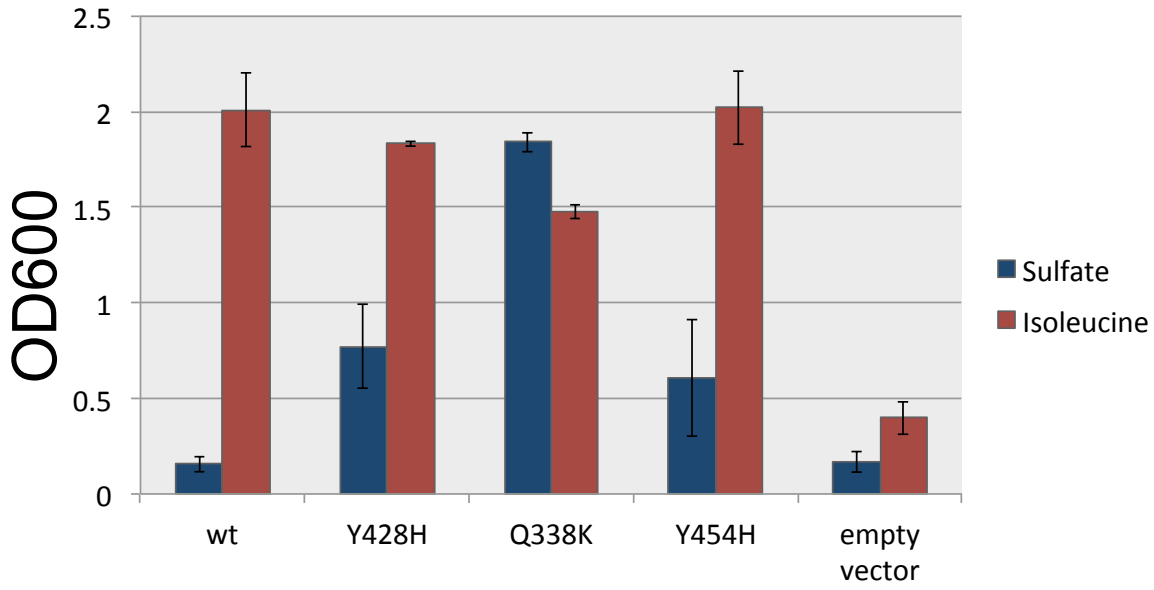


Figure 4.5 *YIL166C* encodes a broad specificity transporter. Allele replacement strains (*sul1*Δ *sul2*Δ *yil166c*Δ) expressing evolved alleles and the wt allele on pRS413 were grown in the following conditions: **(Dark Blue)** Saturation densities for cultures grown in (3mg/L ammonium sulfate) sulfate-limited show that evolved *YIL166C* alleles enable growth and thus contain efficient sulfate transporters. (**Dark Red**) Saturation densities for cultures grown in –N media supplemented with isoleucine as the sole nitrogen source show that expressing *YIL166C* alleles is beneficial relative to the empty vector control. Additionally, all alleles have similar growth except $YIL166C^{Q338K}$, which shows a significant reduction in growth relative to wt *YIL166C*, $YIL166C^{Y428H}$, and $YIL166C^{Y454H}$.

Phenotypic screening of *yil166c*Δ and wt suggest that *YIL166C* encodes a broad-specificity transporter of nitrogen and sulfate sources.

YIL166C is one of 184 MFS domain-containing proteins in yeast (Yan 2013). These domains are responsible for primary and secondary transport of diverse substrates including

carbon, nitrogen, and phosphate sources. To better understand the range of substrates transported by *YIL166C* we compared the growth of wt (FY4) to *yil166c*Δ yeast using the Omnilog platform to assay 768 different media conditions in duplicate. From these data we calculated doubling times and found that *yil166c*Δ strains grew significantly more slowly in media in which specific amino acids (isoleucine and methionine) and a few sulfate sources (butane sulfonic acid and isethionate) served as sole nitrogen or sulfur sources respectively.

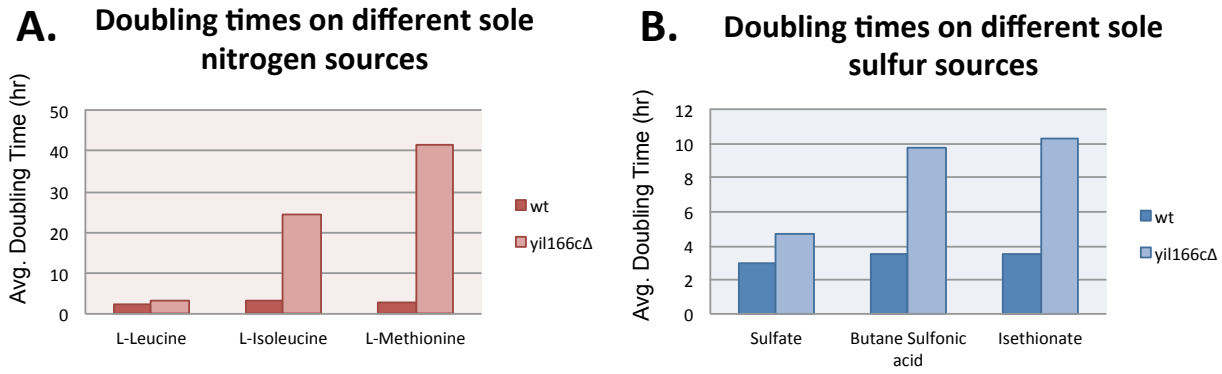
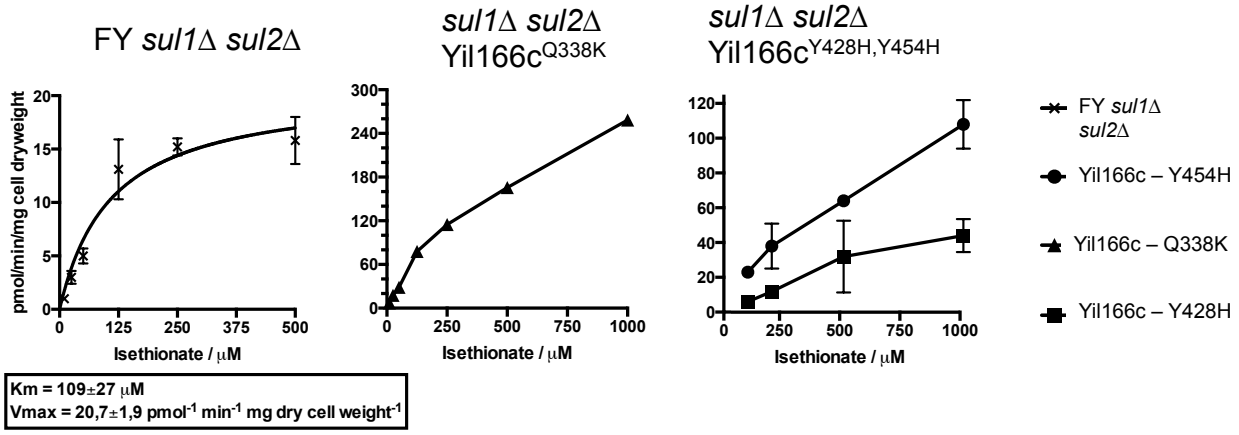


Fig 4.6A,B. These averaged doubling time data in replicates suggest that *YIL166C* encodes a broad-specificity transporter capable of transporting both **A.** nitrogen (leucine, isoleucine, and methionine in minimal nitrogen-free media) and **B.** sulfur (sulfate, isethionate and butane sulfonic acid) sources.

Pleiotropy in affinity for evolved sulfate-transporting *YIL166C* alleles.

To better characterize the kinetics of binding and uptake for sulfur sources I (to a lesser extent) and a collaborator Sylvester Holt (to a greater extent) from Johan Thevelein's lab at KU Leuven in Belgium have performed radionuclide update assays for uptake of sulfate and isethionate on evolved clones. Our data (**Fig 4.7A,B.**) show that ancestral *YIL166C* has a high affinity for isethionate (K_m = 109μM).

A. Uptake of ^{35}S Isethionate



B. Transport of $^{35}SO_4$

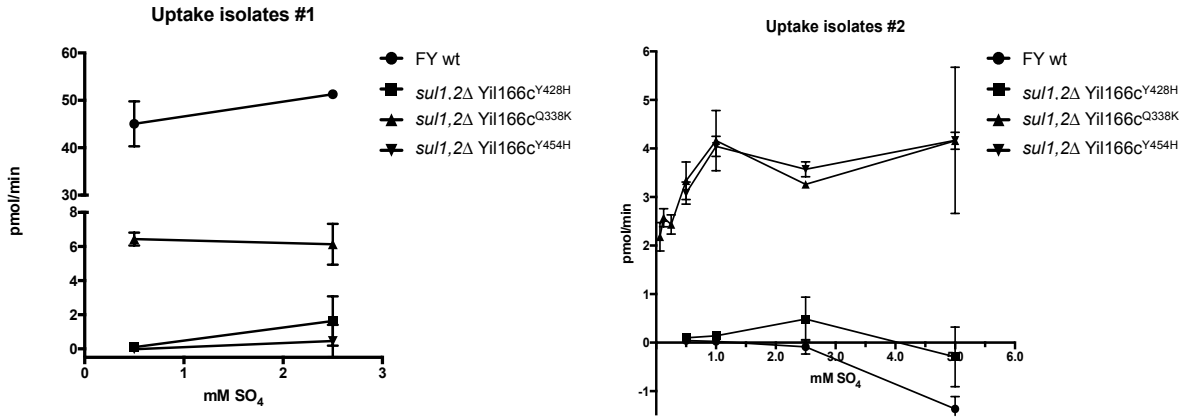


Figure 7A,B. ^{35}S –uptake assays to test each evolved clone (A1,B1,C1) as well as wt and *sul1*Δ *sul2*Δ ancestral strain affinity for **A.** isethionate and **B.** sulfate.

These data also show (**Fig 4.7B.**) that *YIL166C* is initially a low affinity sulfate transporter that is converted to a high-affinity sulfate transporter to compensate for the loss of *SUL1* and *SUL2*. Testing of the allele replacement strains that I generated to determine which alleles are expressed in the context of a *sul1*Δ *sul2*Δ *yil166c*Δ background, is currently under way. It is expected that these assays will help more directly quantify the changes in affinity conferred solely through expression of the alleles and not indirectly along with other mutations in the particular background. However, even without sulfate uptake data for the allele replacement strain, our data

from the evolved clones suggest that the *YIL166C*^{Q338K} allele has the strongest affinity for sulfur sources and the weakest affinity for nitrogen sources (**Fig 4.6A,B.)**.

4.5 DISCUSSION

Amplification of *SUL1* under sulfate limitation is a well-documented finding that emphasizes yeast's reliance on this protein for adaptation to sulfate limitation. Upon removal of this protein through gene deletion we see convergent compensatory evolution in which *SUL2,* the paralog of *SUL1,* is amplified in 4/4 evolved populations subjected to array CGH. While this genotyping method is insufficient to reveal fine scale mutations like SNVs and small indels, we observe a clear bias towards mutation through 'amplification' of a protein identified to have a similar transport function to compensate for the lack of the preferred transport protein. It is striking to observe such a strong bias towards amplification of *SUL1* and never *SUL2* in the wt FY strains of haploid yeast. Based on the repeatability and importance of sulfate transport relative to other targets of adaptation in other nutrient environments (**Chapter 3**) we reasoned that if we were to grow a strain lacking both known sulfate transporters (*sul1*Δ *sul2*Δ) for prolonged periods of time at incrementally lower concentrations of sulfate, we would require the strain to acquire a new or previously uncharacterized method of sulfate transport. After two months of growth and roughly ~170 generations we identified 3 populations that could grow on concentrations of sulfate known to be growth-limiting for wt yeast. Whole genome sequencing of clones from all three sulfate transport-proficient strains identified likely adaptive mutations in *YIL166C*.

Although there is no protein structure of Yil166c or any close homologues, a number of lines of evidence support the supposition that the evolved alleles might influence substrate affinity or preference. For example, all three evolved mutations *YIL166C*^{Q338K}, *YIL166C*^{Y428H}, and

YIL166C^{Y454H} are located in putative trans-membrane domains and result in changes in amino acid charge, both of which are common characteristics of mutations that change either the affinity or specificity of a transporter protein. However, the observation that single-nucleotide mutations turned (what appears to be) a broad specificity transporter capable of efficient isethionate transport into an efficient sulfate transporter provides valuable information for future structure-function analysis.

Informatics efforts to determine the likely substrate of MFS domain-containing proteins in yeast have suggested that *YIL166C* is predicted to be an allantoate transporter. Informatic predictions must be refined over time as empirical data are produced. We present phenotypic growth rate data from a range of media conditions that demonstrate that *YIL166C* is likely to encode a broad specificity transporter. Loss of *YIL166C* yielded reduced growth rate on both alternate nitrogen sources (such as isoleucine) and sulfur sources (such as sulfate and isethionate). Evidence that *YIL166C* may play a role in sulfate utilization has also been reported through early studies of physiology in chemostats where it was shown that growth of wt yeast in sulfate-limited media results in greatly increased expression of *YIL166C* (Boer, de Winde et al. 2003).

In our current study this physiological evidence is supported by multiple observations. Firstly, convergent evolution for mutations in *YIL166C* in 3/3 evolved populations suggests that *YIL166C* is the next most important target of mutation in the absence of *SUL1* and *SUL2* in sulfate-limited environments. Secondly, we demonstrated through tetrad analysis and allele replacement experiments that evolved alleles of *YIL166C*^{Y428H}, *YIL166C*^{Q338K}, and *YIL166C*^{Y454H} were sufficient for growth on sulfate-limited media, thus implicating them as efficient transporters of sulfate. Thirdly, sulfate uptake assays of evolved clones demonstrated that

evolved *YIL166C* alleles are capable of efficient uptake of sulfate. However, sulfate does not appear to be the substrate for which *YIL166C* has greatest affinity. Evolved alleles of *YIL166C* with improved affinity for sulfate showed a pleiotropic and concordant increase in ability to transport isethionate. Limited experiments testing the growth rates of wt vs *yil166c*Δ (**Fig 4.6A,B**) or saturation densities of wt and evolved alleles of *YIL166C* (**Fig 4.5**) suggest that *YIL166C* is likely a broad specificity transporter that transports nitrogen sources including isoleucine. The fact that the best sulfate transporting evolved allele $YIL166C^{Q338K}$ had reduced growth on isoleucine as sole nitrogen source suggests an antagonistic pleiotropic relationship between import of these two nutrient types for this protein.

In summary, the evolved alleles of $YIL166C^{Q338K}$, $YIL166C^{Y428H}$, and $YIL166C^{Y454H}$ all show pleiotropic improvements in transport of sulfate as well as isethionate. We see convergent compensatory evolution in which an apparent broad specificity transporter may be repurposed to compensate for the loss of sulfate transport proficiency. Our results highlight the possibility that broad-specificity transporters might be the substrate of compensatory evolution for loss of canonical transport function. Additionally, this work is an example of how we may be able to use compensatory experimental evolution to reveal proteins capable of acquiring new function and thus better understand the potential for novelty in yeast evolution.

4.6 METHODS

All knockout strains were generated in a haploid FY background initially auxotrophic for *URA3* in both mating types. Unless otherwise stated the resulting strains were prototrophic as each knocked out gene was replaced with *Kl-URA3.* Subsequent mating of *sul1*Δ and *sul2*Δ strains was used to generate a *sul1*Δ *sul2*Δ double knockout strain. A haploid *MATa sul1*Δ strain was

used for chemostat evolutions and a haploid *MATa sul1*Δ *sul2*Δ strain was used for subsequent serial batch-transfer experiment.

To better understand the contribution of the uncharacterized gene *YIL166C* and mutations therein to adaptation and physiology I generated a *yil166c*Δ strain using the method described above for *sul*-knockouts. This *MATa* haploid strain was subsequently used for the Omnilog phenotypic screen described below. However it was discovered that this strain contained unintended mutations that prevented growth on galactose. Follow-up experiments involving backcrossing or sequencing of this strain did not prove informative as to what these background mutations might be and thus I re-made the strain in duplicate; both resultant prototrophic *yil166c*Δ strains did not show any inability to grow on galactose. To generate strains suitable for allele replacement I crossed one of these new haploid *yil166c*Δ strains to the aforementioned *sul1*Δ *sul2*Δ ancestral strain for use in the batch sulfate evolution experiment. Spores from this cross were then crossed to an FY his auxotroph (*his3*). Two spores derived from this cross had the desired *sul1*Δ *sul2*Δ *yil166c*Δ *his3* genotype and were saved for further allele replacement experimentation such as that described in (**Fig 4.5**).

Plasmids

To better understand the specific contribution of each evolved allele to nutrient uptake phenotypes, wt and evolved alleles of *YIL166C* were cloned into pRS413 using standard Gibson-cloning techniques, were validated via Sanger sequencing, and transformed into the his-auxotrophic *sul1*Δ *sul2*Δ *yil166c*Δ strain described above.

Illumina Sequencing of Evolved Sulfate Transport Proficient Clones

DNA samples from evolved clones and populations were prepared for WGS using Illumina Nextera kits according to the provided protocol. Libraries were sequenced on an Illumina NextSeq. The resultant fold coverage is represented in (**Fig 4.3B.**). Reads were aligned with BWA (Li and Durbin 2009) and SNVs were called using Samtools (Li, Handsaker et al. 2009) after applying standard filters.

Omnilog growth phenotype screen of wt and *yil166c*Δ yeast.

Cells were grown in 768 different media conditions that assayed growth and metabolic activity on a range of carbon, nitrogen, sulfur, and phosphate containing molecules as sole source of each given element. Additionally this panel of media conditions tested a range of salt, pH, and toxic compounds such as antibiotics and known metabolic poisons and DNA damaging reagents. Use of a metabolic dye, tetrazolium, that is reduced to a purple compound, formazan, was included to quantitate cell growth in each well. Growth curves were photographic quantification of formazan production. We then calculated doubling times based on log-phase growth data for each well, which was then averaged across replicates.

Kinetic (Radionuclide uptake) assays

Evolved clones were grown on SC-HIS in a fresh culture to OD 1, and subsequently starved for sulfur 3 days in sulfur starvation media. Then starved cells were exposed to a range of $^{35}SO_4$ and ^{35}S-Isethionate 2,5 to find the kinetic constants for each evolved allele of *YIL166C*. Work to characterize the allele replacement strains described above is currently under way and should provide better direct measurement of the extent to which each allele in isolation contributes to sulfate and isethionate uptake.

4.7 FUNDING ACKNOWLEDGEMENTS

AWM was funded by both the (GTG) genome training grant and IDTG interdisciplinary training grant in cancer co-administered by UW and the Fred Hutchinson Cancer Research Center.

Chapter 5. MULTIPLEXED MUTATION RATE ASSAYS

5.1 ACKNOWLEDGEMENTS

The following work started as a collaboration with Alison Gammie and makes use of the alleles she created for her 2007 Genetics paper (Gammie, Erdeniz et al. 2007). The initial proof of concept experiments described in this chapter are a work in progress and are being followed up by a rotation student, Jolie Carlisle. We are currently producing the requisite data to generate a paper describing this method. Maitreya and I conceived of all experiments. Initial measurements of mutation rates in wt and *msh2*Δ strains were performed by Maitreya and repeated again more recently by Jolie Carlisle under my supervision as part of her rotation project. Otherwise, I performed the analysis and made all figures presented in this chapter.

This chapter is part of a manuscript in preparation.

5.2 ABSTRACT

Cancer is a disease of genetic alterations. In particular, single nucleotide variants (SNVs) present as either germline or somatic point mutations are essential drivers of tumorigenesis and cellular proliferation in many human cancer types. Contemporary sequencing surveys of tumor transcriptomes, exomes, and genomes have begun to reveal mutations in these commonly affected cancer genes in addition to many novel mutations in genes with no previous implication in cancer. This unprecedented opportunity for SNV discovery in cancer also creates unmet need to figure out which SNVs lead to pathogenesis and the mechanism through which these mutations operate.

Current methods of predicting the functional impact of SNVs based on conservation are insufficient for accurate quality diagnosis and treatment (Liu, Jian et al. 2011). Traditional means of allele phenotyping and mechanism determination in model systems such as yeast, worms, mice, and cell culture are not nearly able to keep up with the rate of SNV discovery and will continue to grow further behind each year. Thus, it is essential to develop new tools to speed functional and mechanistic studies in model systems for the medical community to transcend trial and error based ascertainment of the disease risk conferred by a given allele. In this chapter I describe my efforts, with the help of others, to come up with a quick and accurate approach to measure mutation rate.

5.3 INTRODUCTION

Mismatch repair in cancer and tumorigenesis.

Colorectal cancer is a leading cause of mortality in the United States. It is estimated that 2012 saw 143,000 new diagnoses and 72,000 deaths from colorectal cancer in the United States (Siegel, Naishadham et al. 2012). Lynch syndrome, also known as hereditary non-polyposis colorectal cancer (HNPCC), is a highly penetrant autosomal dominant syndrome caused by defects in mismatch repair genes that accounts for about 4 to 6% of colorectal cancer cases each year (Lynch, Lanspa et al. 1988, Lynch and Lynch 1993). Lynch syndrome is typified by early onset colorectal cancer with an average age of onset of 42 as compared with 72 for onset of colorectal cancer in the general population (Mecklin and Jarvinen 1991, Lynch and Lynch 1993). DNA mismatches form primarily during replication of the genome (Kunkel and Erie 2005). Defects in the *MSH2* gene may account for as many as 50% of Lynch syndrome cases, and along with defects in *MLH1* or *MSH6* likely account for the vast majority of Lynch syndrome cases (de la Chapelle 2005).

In the past family history played a predominant role in diagnosis of Lynch syndrome as patients were measured against Amsterdam II (Lipton, Johnson et al. 2004) or other criteria for disease risk before proceeding to molecular characterization of disease risk. Once identified as at risk under these criteria patients are then tested for microsatellite instability, a hallmark of mismatch repair defective genomes, or immunohistochemistry is used to test for decreased expression of mismatch repair proteins often associated with defective alleles. Each of these testing methods used as selection criteria for genotyping are imperfect at identifying disease risk. One recent study has demonstrated that the sensitivity of the Amsterdam II criteria is between 38-87% for identifying individuals at risk for Lynch Syndrome based on which mismatch repair

gene had a germline pathogenic variant (Sjursen, Haukanes et al. 2010). As such clinics have begun genotyping patients for *MSH2*, *MSH6*, and *MLH1* if they show any characteristics associated with, or are related to someone who has colorectal cancer. Although genotyping is currently the gold standard for determining disease risk, pathogenicity is known for only a small subset of SNVs. Even clinically-associated SNVs can be spuriously cataloged. For instance, it has been reported that G322D and L390F are common missense mutations in *MSH2* that have no association with Lynch syndrome and yet their pathogenic classification persists in the databases based on the fact that these mutations were found in Lynch syndrome patients (Gammie, Erdeniz et al. 2007).

The need for model systems to study the effect of variation on mismatch repair genes in humans.

The goal of understanding the pathogenic effect of variation in mismatch repair genes cannot currently be comprehensively assayed in mammalian systems. We lack the methods to stably express or integrate thousands of alleles, measure mutation rate, and link genotypic variation to phenotypic variation in a rapid and accurate fashion. However, model systems benefit from identical mismatch repair mechanisms mediated by highly conserved homologs in this ancient pathway. Much of the foundational mismatch repair research was accomplished in model organisms. Basic science investigations conducted in bacteria and yeast were essential for the characterization of genes involved in mismatch repair and ultimately the discovery of an array of clinically relevant mutations in DNA mismatch repair genes (Peltomaki and Vasen 2004). Numerous studies have already used mismatch repair in yeast, where gene function can be measured in mutation rate assays to identify pathogenic alleles in humans (Peltomaki and Vasen

2004, Gammie, Erdeniz et al. 2007, Takahashi, Shimodaira et al. 2007, Arlow, Scott et al. 2013). That having been said, no methods used to date are multiplexed and thus suited to deal with the ever-growing list of missense alleles or other mutations we might want to test for effect on mutation rate.

Chemostats are an excellent tool to quantify mutation rate (Novick and Szilard 1950) as at steady-state growth they are free from the extreme fluctuations of batch cultures described by Luria and Delbruck (Luria and Delbruck 1943). During each doubling of the chemostat culture, half of the cells are diluted out, leading to a linear accumulation of neutral mutations at the mutation rate. Therefore, the mutation rate can be directly measured by calculating the slope of the mutation frequency over many generations of growth. The primary alternative to this method is the fluctuation assay (Luria and Delbruck 1943), which requires many replicates to account for the noise generated by jackpot effects seen in batch cultures. Fluctuation assays can be used to measure mutation rate only for one allele at a time.

5.4 RESULTS

I have prototyped this method on a library of *msh2* variants found in colorectal cancer patients. This library had already been generated and characterized in a *msh2Δ* background individually for mutation rate and other phenotypes (Gammie, Erdeniz et al. 2007). In this work, all missense alleles were characterized primarily by using a qualitative 'patch' assay to gauge mutagenicity of each allele relative to wt and null alleles. In this way, wt alleles are likely to show no growth on canavanine and *msh2*Δ strains are likely to form numerous micro-colonies.

To better quantify more alleles at once, and to do so more accurately, we turned to the chemostat approach. To start we determined mutation rate for wt and *msh2Δ* in haploid yeast under glucose-limited growth. For this experiment the two strains were grown in separate chemostats for the duration of the experiment. In chemostats a *msh2Δ* strain shows a 20-fold increased mutation rate relative to wt, similar to estimates in the literature.

To see if mutation rate for all 50 of these variants could be determined simultaneously we co-cultured all 50 variants in a chemostat for 35 generations. The population was sampled roughly every 5 generations and 15,000 colonies were plated to both complete media lacking histidine (C -His) and containing canavanine (30mg/L Can in C -His –Arg -Ser media), which selects for mutagenic alleles (see **Fig 5.1A.**). After two days growth colonies were scraped, plasmid DNA was extracted and a Nextera (Next-Gen) sequencing library was made from the amplified and gel-extracted *MSH2* locus from the pool of alleles for each sample. The slope of the proportion of a given allele on canavanine-containing media relative to C-His over time estimates the mutation rate for each of the alleles (see **Fig 5.1B.**). Custom python scripts were used to determine the frequency of each allele by counting the missense mutations that uniquely identify each allele relative to the reference sequence contributed by other alleles in the pool.

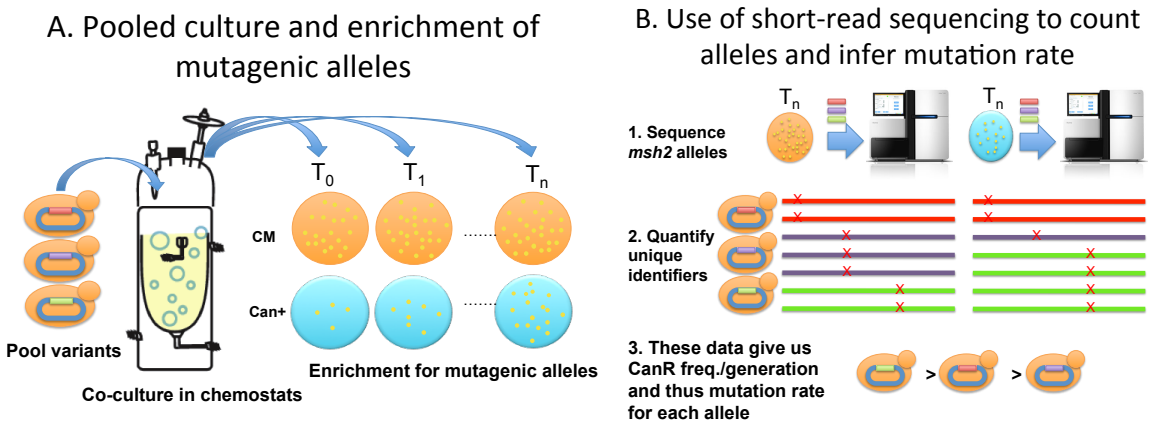


Figure 5.1A,B. Multiplexed mutation rate analysis in chemostats. A. Pooled variants may be co-cultured in chemostats and plated to media containing canavanine to enrich for mutagenic alleles. Plating cells to media without canavanine can be used for normalization. B. *MSH2* loci are amplified, sequenced, and used to infer CanR frequency/generation conferred by each allele which is used to infer mutation rate.

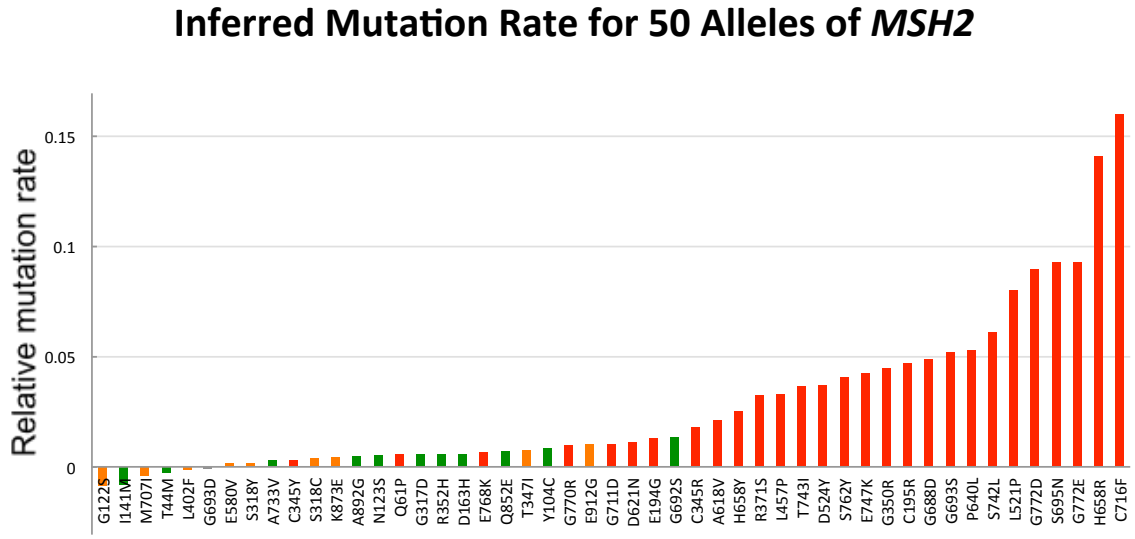


Fig 5.2. Multiplexed characterization of mutation rate for 50 alleles of *MSH2* simultaneously. (color coding matches patch assay results from *(Gammie, Erdeniz et al. 2007)* where Red = pathogenic, Orange = intermediate, Green = wt-like mutation rates)

These data (**Fig 5.2.**) are promising because they recapitulate our expectation that alleles that confer high-mutation rate in the patch assays appear to have higher relative mutation rates in our multiplexed mutation rate assay.

5.5 Discussion and ongoing experiments

While these preliminary results are exciting and suggest that we should be able to screen hundreds or thousands of alleles at once, more work is needed to validate the accuracy of this approach and better estimate the scale at which it can be used. Experiments are currently underway to quantify mutation rate for 6 different alleles of *MSH2* with a range of mutation rates, in triplicate, screened one at a time in separate chemostats. Additionally, a rotation student, Jolie Carlisle, and I have recently performed another three pooled allele experiments with the same 50 alleles of *MSH2*. By comparing the pooled experiments amongst each other we will quantitate how precise multiplexed mutation rate analysis is. By comparing to mutation rates determined for each allele separately and in triplicate we will be able to determine how accurate pooled methods are vs. the one at a time approach. Should these pooled methods prove both precise and accurate for mutation rate analysis we are well poised to try them on more alleles of *MSH2* as I have already generated 100 barcoded alleles of *MSH2* via overlap extension PCR. Many of these were designed so that we could test *msh2* variants of unknown significance, and thereby begin to better understand Lynch syndrome risk in humans. Ultimately, the goal of this approach is not to study variation only in *MSH2* but also in a variety of other mismatch repair genes or genes that confer variation in mutation rate. Toward this end I have begun to generate alleles of *MSH6* and *MLH1,* which I intend to test should our method of measuring mutation rate be validated.

Plasmid and strain generation

Plasmids used in (**Fig 5.2.**) were generously provided by Alison Gammie (Gammie, Erdeniz et al. 2007). As were the strains used for initial iterations of this experiment. It was found that this mutator strain could bypass the auxotrophic markers present in the strain through reversion, which called into question how useful it would be for our pooled screening approach in chemostats. Thus I generated a new strain by crossing an *msh2*Δ prototroph to the *his3*Δ strain from the deletion collection, which carried with it a variety of auxotrophic markers. From this cross I isolated a spore with the genotype *msh2*Δ *his3*Δ and carried it forward for experimentation.

Transformation and cell culture in chemostats

Competent *msh2*Δ cells were transformed with pRS413 plasmids containing 50 different missense alleles of *MSH2* provided by Alison Gammie. These cells were selected for incorporation of the vector through plating to C-His media. In total >10,000 transformants were collected, pooled, and frozen in concentrated glycerol for use in pooled chemostat experiments. $1/10^{th}$ of the cells collected were used for inoculation of chemostat experiments. Pumps were turned on directly after inoculation and the population was sampled roughly every 5 generations. 15,000 colonies were plated to both complete media lacking histidine (C -His) and containing canavanine (30mg/L Can in C -His –Arg -Ser media), which selects for mutagenic alleles. After two days growth colonies were scraped, plasmid DNA was extracted, and a Nextera (Next-Gen) sequencing library was made from the amplified and gel-extracted *MSH2* locus for each of the two media conditions.

Sequencing analysis

Custom Python scripts were used to analyze the raw reads and quantify how often each of the unique identifier sequences were observed vs. reference on Can+ and Can- media. This proportion of alternate vs. reference allele sequences were used to identify the proportion of each allele in the population at each time-point. The slope of the proportion of a given allele on Can+ media relative to Can- over time can be used to estimate the mutation rate for each of the alleles (see **Fig 5.1B.**).

5.6 Funding Acknowledgments

AWM was funded by both the (GTG) genome training grant and IDTG interdisciplinary training grant in cancer co-administered by UW and the Fred Hutchinson Cancer Research Center. For the IDTG I was co-advised by Polly Newcomb.

Chapter 6. CONCLUSIONS AND FUTURE DIRECTIONS

This thesis describes my efforts to generate and validate multiplexed continuous culture arrays (both chemostats and turbidostats). Experiments to better define the spectrum of mutations that arise across dozens of evolution experiments identified 769 mutations across 95 evolution experiments, which will be a useful dataset for those wishing to understand the repeatability of evolutionary trajectories. Repeatability and analysis of gene function for these populations provided unique insights into previously unappreciated and important adaptive strategies such as gene silencing and chromatin remodeling in phosphate limitation. Additionally, repeated mutations in glutathione metabolism (*GSH1)* and acquisition of mutations in bud morphology gene *BEM2* highlight these adaptive strategies as worthy of further study. Lastly, I observed an unprecedented level of single nucleotide variants in *SUL1* and *PHO84*, thus highlighting these as a further means to adapt in addition to the more commonly reported copy number variation of these same genes. However, my work to date has not solved the problem of how to more easily and comprehensively quantify epistasis. Quantification of how epistasis has shaped a particular evolutionary trajectory seen in an evolving population must still be tested one genotype at a time. However, I do have some ideas of how to do this and have even begun to generate some genetic libraries that could be tested (see **Chapter 6.1**). I have applied methods similar to evolutionarily-derived suppressors to identify a novel low affinity sulfate transporter. This approach, of completely removing all genes of a particular function, is a useful and underutilized approach to identify novel proteins that compensate for that same function through adaptive evolution. This work has led me to a new perspective about the nature of experimental evolution (**Chapter 6.3**) and has led to some specific applications. For instance, a recent effort to find suppressors of the reciprocal sign epistasis between *mth1* and *HXT6,7* found no such suppressor, and thus likened

this area of the fitness landscape to 'Death Valley'. My studies with a *sul1*Δ *sul2*Δ strain suggest a more direct approach may be more efficacious (**Chapter 6.2**). Lastly, I have worked on developing a multiplexed mutation rate assay and prototyped its use to study alleles of mismatch repair gene *MSH2*; however, I have also come up with strategies to use this general approach to test combinatorial variation between alleles of mismatch repair genes relevant to Lynch syndrome (**Chapter 6.4**).

6.1 How can we more comprehensively map fitness landscapes genome-wide?

WGS, of individual evolved clones and their ancestors, followed by detailed physiological studies, allelic replacement experiments, and fitness assays has dramatically improved our understanding of the genotype-phenotype map. However, these approaches alone are labor-intensive and thus rarely used to create detailed maps of adaptive landscapes. How this can be improved by greatly increased replicate number and genetic screens are central to this thesis.

It seems as though we are moving towards a day when fitness landscapes will not be some abstract smooth surface, devoid of actual genotypes or measures of fitness, and instead they will be infused with a multitude of real data. Indeed, massively parallel methods to functionally characterize proteins have been employed to do just this. Analogous genome-wide methods are still lacking across taxa. However, recent work in our lab by Celia Payen and others has exploited the deletion collection, as well as the CEN-MoBY (Ho, Magtanong et al. 2009) and 2μ (Jones, Stalker et al. 2008) collections, to test single gene copy number increases and decreases to mimic gain and loss of function mutations (Payen, Sunshine et al. 2015). By testing these collections in commonly imposed nutrient limitation regimes in chemostats we may better

be able to rationalize which mutations observed in evolution experiments in specific environments are likely to be adaptive. I discuss this in greater detail in (**Chapter 3**), but as a useful example this study found that increased expression of *SUL1* was most beneficial (~42% increased fitness) in sulfate limitation. This finding is consistent with the virtually universal amplification of *SUL1*. However, altered copy number of *MAC1* and *PHO3* were identified in this study as beneficial (conferred a 25% and 31% increased in fitness, respectively) in sulfate limitation and are not seen in evolution experiments to date. It is possible that these findings are artifacts of the screening method, caused by unintended background mutations, or that these ORFs reside in regions of the genome with reduced mutability and as a consequence are not seen in evolution experiments. On the other hand, it is possible that early large effect mutations such as *SUL1* amplification that often begin to sweep the population after 50 generations of growth (Payen, Di Rienzi et al. 2014) may constitute an altered genetic background where mutations screened in the context of a 'wt' background are no longer beneficial. One interesting candidate for such an interaction is that increased expression of *PHO3* is greatly beneficial in the 'wt' background but not observed in yeast populations evolved in sulfate limitation, wherein *SUL1* amplification might ostensibly have already occurred. *PHO3* is a constitutively expressed acid phosphatase that hydrolyzes thiamine phosphates in the periplasmic space, increasing cellular thiamine uptake. Thiamine is a sulfur-containing molecule; thus, may provide an alternative means for importation of sulfur, which is less beneficial if *SUL1* is already amplified. Although the media used in chemostat experiments is not supplemented with thiamine, thiamine may be secreted or imparted to the media upon cell lysis, thus providing a beneficial effect on fitness either in a pooled screening scenario, or as a sub-population within an evolving population grown in a chemostat.

To investigate epistasis on more than just a case-by-case basis I have cloned *SUL1* onto a CEN plasmid (pRS413) and intend to transform this into the deletion collection to assay how *SUL1* amplification influences fitness associated with loss of gene function for all knockouts in the collection. I have also generated a strain in which *SUL1* has been knocked-in at the mating type locus to increase copy number to two. This strain can be used as a chassis to screen the CEN-MoBY collection to see how increased copy number of *SUL1* influences the finesses conferred by all ORFs in the CEN-MoBY collection. Taken as a whole, I think methods such as these will provide an interesting avenue for more comprehensively quantitating how each gain or loss of function mutation is likely to affect fitness in a genomic context directly applicable to a given evolving genome in the specific environment in which it adapts.

6.2 Finding a way out of the 'valley of death'.

One of the most striking features of a topological fitness landscape that I have discussed in this thesis is that of reciprocal sign epistasis. One of the few examples of this characterized and observed in the context of experimental evolution is the case where loss of function for *mth1* and amplification of *HXT6,7* - two common beneficial mutations for yeast evolved under glucose limitation - have a combined net-negative effect on fitness. Thus, the two mutations are not seen to co-occur in the same cells although they can co-exist and clonally interfere in populations. Further work by Sherlock and colleagues involved starting their evolution experiments with either a nonfunctional *mth1* allele or *HXT6,7* amplified fixed within the population. They then evolved these strains to see whether the other mutation occurs in the evolution experiment. Since no mutations were found that suppressed the negative fitness effect of the *mth1 HXT6,7* genotype

the authors likened this region of the fitness landscape to a 'Valley of Death' (Chiotti, Kvitek et al. 2014).

Based on my studies of compensatory evolution of *sul1*Δ and *sul1*Δ *sul2*Δ strains under sulfate limitation (**Chapter 4**), I suggest that the simplest way out of the reciprocal sign epistasis associated with loss of *MTH1* function and amplification of *HXT6,7*, would be to evolve the strain containing both mutations under glucose limitation until it a returns to positive fitness; this experiment should identify mutations that are sufficient to suppress the negative fitness of this genotype. In the parlance of Sherlock *et al.* these mutations could be appropriately considered a way out of this 'Valley of Death'.

6.3 Evolution as an ongoing supressor screen

For the compensatory evolution experiments described in (**Chapter 4**), wherein *sul1*Δ *sul2*Δ strains were grown on media containing incrementally lower sulfate concentrations with the intent of finding new or undiscovered sulfate transport mechanisms in yeast, we trod the line between an evolution experiment and a suppressor screen. Our method is distinct from a classical suppressor screen, as a number of mutations are allowed to accumulate and collectively allow for the acquisition of the desired phenotype – efficient sulfate transport. Although this method is more of an evolutionary process than a simple suppressor screen – the semantics could cut both ways – perhaps the deeper nature of this discussion is that evolution is in and of itself an ongoing suppressor screen where a non-optimized organism continually suppresses the lack of optimization through sequential accumulation of beneficial mutations over time.

6.4 TOWARD COMBINITORIAL SEQUENCING STRATAGIES TO CHARACTERIZE TWO-LOCUS EPISTASIS.

Epistasis is not just important because of its role in adaptive evolution, but is also central to understanding disease risk. For many cancer syndromes, including Lynch Syndrome, which is caused by mutations in *MSH2*, *MSH6*, *MLH1*, *PMS1*, and others, disease severity is modulated by variation in two or more mismatch genes within an individual. Cases have been documented where variation in *MSH2* and *MSH6,* which are individually tolerated, result in Lynch syndrome when both are present in a patient (Kariola, Otway et al. 2003). Similar findings focusing on different alleles have also been identified using one-by-one testing of 8 combinations of alleles of *MSH2* with alleles of *MSH6*. My analysis of recent exome sequencing data (Tennessen, Bigham et al. 2012) for individuals not thought to have Lynch syndrome suggests that having variation in both *MSH2* and *MSH6* is not uncommon. Of 2,500 exomes analyzed, 16.4% of individuals had missense mutations in *MSH2* and roughly half of those individuals (7.5% of the total) also had missense variation in *MSH6*. Development of multiplexed methods to determine the effect of epistasis is necessary to understand complex diseases like cancer, and timely given the availability of ever-growing genotypic information for patients and unaffected individuals alike.

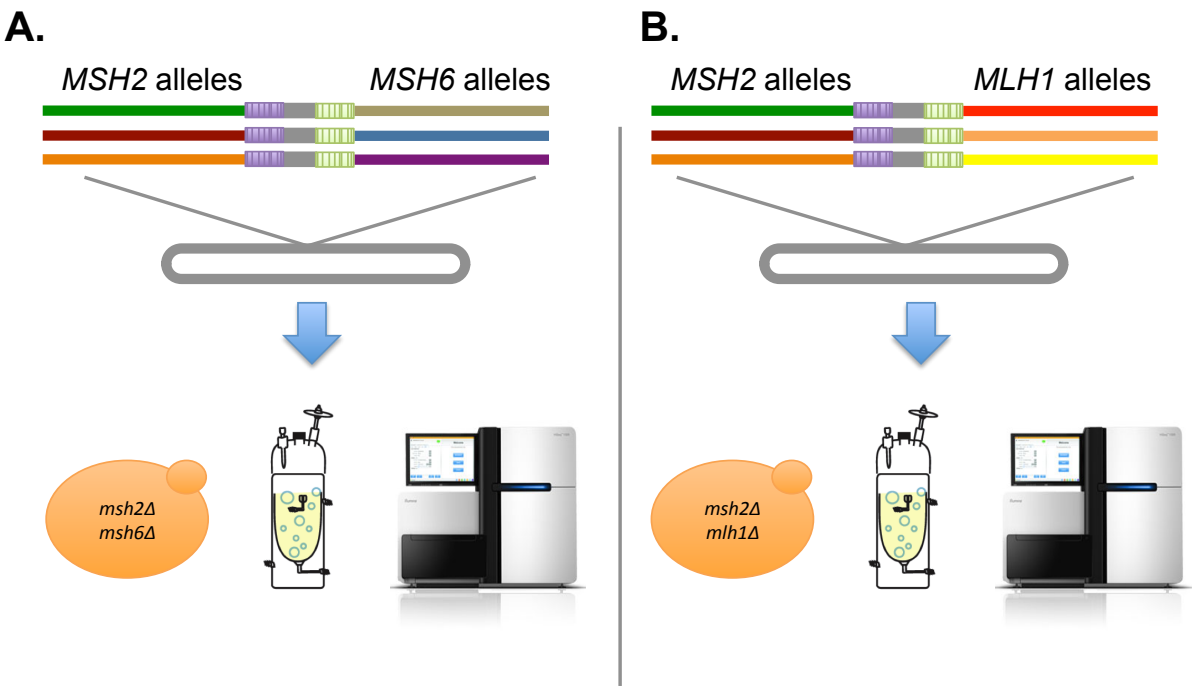


Figure 6.1. Multiplexed characterization of epistasis between barcoded alleles applied to A. mismatch repair genes *MSH2* and *MSH6*. B. Pairwise combinatorial variation between alleles of *MSH2* and *MLH1*.

To study simultaneous variation between barcoded alleles of *MSH2* I have created a hundredof alleles of *MSH2* and tens of alleles of *MSH6* in such a way that they can be PCR-fused together before cloning. With this cloning scheme (see **Fig 6.1**) the barcodes on the 3' end that uniquely identify each allele will be cloned adjacent to each other. In this way pairs of alleles can be screened simultaneously and sequencing of the two barcodes in a single read will link the specific alleles to the inferred mutation rate (**Fig 6.1A.**). This work could be used to study pairwise variation in a wide variety of contexts, including variation between alleles of *MSH2* and *MLH1* (**Fig 6.1B.**), as variation in these two proteins are thought to be responsible for roughly 95% of Lynch syndrome cases.

Final thoughts:

There are many questions that remain about evolutionary dynamics and the spectrum of mutations responsible for adaptation. Experimental evolution is by its very nature labor-intensive, and for evolution experiments excess mutations are usually uncovered, so much so that it would be hard to functionally test all of them one at a time. As such, I am excited by genomic screening techniques that may one day facilitate more comprehensive functional characterization of genotypes that arise in evolution experiments. CRISPR-based methods to regulate gene expression (Gilbert, Larson et al. 2013) may prove an attractive approach. For example, CRISPR can be used to decrease expression for multiple genes at once, which may aid in the inference of fitness effects associated with multiple LOF mutations often seen in evolution experiments. The idea to use genetic competitions to better predict evolutionary outcomes has merit. In refining the screening approach used, we may in-turn learn more about the forces that shape and constrain evolutionary trajectories. Additionally the ability to inactivate many genes at once might allow us to determine the extent to which aspects of genetic background, such as redundancy in gene or pathway, influence adaptive trajectories.

Evolution in nature is ultimately thought to involve dynamic environments and heterogeneous selections. As researchers grow more knowledgeable about adaptation in constant environments I would be interested to know how changing environments and extreme environments act on genomic evolution. Devices like the turbidostats that I worked on with Chris Takahashi of Eric Klavins' lab are one example of devices that can change environments; they are modular in nature and can mix multiple media types containing diverse substrates. Additionally, I am interested to see how control of population size, mutation rate, strength of selection, and control of population dynamics might be used to accelerate the evolution of desired traits.

Appendix A: A flexible open-source turbidostat platform for synthetic circuit characterization

Acknowledgements:

This appendix presents the rationale for developing an open-source multiplexed and flexible turbidostat derived from a paper and a project that I contributed to in collaboration with Chris Takahashi from Eric Klavins lab. As Chris was already a proficient engineer, control theorist, and synthetic biologist, my role on the project was to expedite the development of turbidostats partly through adaptation of the miniature chemostat design described in (**Chapter 2**). Additionally, I acted as a beta tester of many early iterations of the turbidostats we designed; yeast genetics and microbiology consultant. On rare occasions I also helped out with next-gen sequencing of yeast and bacteria cultured in turbidostats. Since I made significant contributions to this project and find it to be a promising device for a multitude of applications, I provide the following modified excerpt from our paper, for which Chris Takahashi is lead author, on the turbidostat designs that we published in *ACS synthetic biology* (Takahashi, Miller et al. 2015). Supplemental data may be found here: http://pubs.acs.org/doi/ipdf/10.1021/sb500165g

Abstract

Engineered biological circuits are often disturbed by a variety of environmental factors. In batch culture, where the majority of synthetic circuit characterization occurs, environmental conditions vary as the culture matures. Turbidostats are powerful characterization tools that provide static culture environments; however, they are often expensive, especially when purchased in custom configurations, and are difficult to design and construct in a lab. Here, we present a low cost, open source multiplexed turbidostat that can be manufactured and used with minimal experience in electrical or software engineering. We demonstrate the utility of this system to profile synthetic circuit behavior in *S. cerevisiae*. We also demonstrate the flexibility of the design by showing that a fluorometer can be easily integrated.

Introduction

The characterization of biological building blocks and complex synthetic systems is fundamental to synthetic biology.(1, 2) Historically, characterization of biological systems has been carried out using batch culture methods.(3-9) The low cost, ease of culture propagation, and scalability through use of well plates and culture flasks have contributed to the prevalence of this culture method. However, in batch culture, the chemical environment is continually changing as cells grow, divide, consume nutrients, and excrete waste products.(10) These changes associated with growth have substantial effects on cellular physiology(11-13) and engineered biological components, which can in turn cause the mischaracterization of a system's input-output response or require the consideration of uncontrolled disturbances to the system.

To more accurately characterize biological systems, researchers must turn to static environments, which yield lower noise in quantitative phenotyping. A primary tool for culturing

cells in a static environment is continuous culture, where inoculated growth medium is continually diluted with fresh medium. At steady state, a continuous culture device will dilute cells and waste products at the same rate that they are being produced leading to an unchanging environment.(14, 15)

Two main categories of continuous culture devices are chemostats and turbidostats. Chemostats dilute liquid culture at a fixed rate and reach steady state when a limiting nutrient has been depleted, or toxins have been accumulated, at which point the growth rate is equal to the dilution rate. Therefore, chemostats are ideal for interrogating the effects of nutrient limitation or where defined growth rates are desired. In contrast, turbidostats use a feedback control loop to keep cell density constant; in a turbidostat, the cell density is continually monitored and used to compute the dilution rate that achieves a prespecified cell density. Thus, turbidostats are ideal for characterization of systems without nutrient limitation, when cells are growing at their maximum rate.

In spite of the advantages, use of chemostats and turbidostats is not commonplace. Factors that limit the use of continuous culture systems include the fact that commercially available bioreactors either lack feedback capability, have such large volumes that dilution rates needed to sustain a log phase culture would be impractical, or are prohibitively expensive. As a result, researchers often design their own turbidostats, which vary widely in implementation based on the needs of the experimenter.(16-23) Unfortunately, the design and construction can be laborious and uncertain. Design decisions that are seemingly innocuous can lead to failures or inflexible devices. Furthermore, development of turbidostats requires specialized knowledge of electrical and computer engineering, which has further limited their popularity.

We have developed an open source turbidostat design (the Flexostat) that contrasts with other publicly available continuous culture devices.(18, 20, 24) Using common hand tools and standard laboratory equipment combined with inexpensive online or university provided fabrication services and 3D printing, our instrument can be built for under $2000 USD. Indeed, many individuals from a wide range of backgrounds have already built turbidostats from our design. Currently, five laboratories have at least one complete instrument, some with custom modifications. As part of our commitment to the community, the authors will continue to contribute updated designs to the Web site and ask, but do not require, that all derived designs be submitted so that the community may benefit.

The Flexostat implements eight parallel culture chambers, which enable users to gather replicate data and perform parallel experiments requiring no extra pumps or controllers. Each chamber contains electronics that enable distributed sensing and mixing control allowing users to add new chamber modules with new capabilities with minimal hardware redesign. A platform agnostic Python program collects data from each chamber and computes dilution rates through a user programmable function.

To demonstrate the utility of our device over standard batch culture, we recharacterized part of the auxin plant hormone pathway cloned into *S. cerevisiae* and grown at various constant cell densities. Previous work has characterized the interactions between the small molecule auxin and the family of F-box proteins in batch culture resulting in a table of key parameters for each interaction.(6) Here, we show that one of these parameters, a degradation rate in the presence of auxin that was previously considered a constant, varies by more than 4-fold depending on the cell density.

Finally, we showcase how the modular design and 3D printed nature of the Flexostat allows us to add new capabilities. In particular, we show a proof of concept single chamber, mixture controlled variant that is capable of detecting strongly fluorescent cultures (the Fluorostat) such as *E. coli* expressing super folder GFP. Fluorescence detection is accomplished through the addition of a single blue excitation source and a pair of low cost, spectrally orthogonal filters. The light sources for the Optical density (OD) measurement and fluorescence measurements alternate in time and share the same photodetector. The dual use of the photosensor allows us to use the same electronics for the standard Flexostat and the Fluorostat, requiring that only software and a single 3D printed part be changed.

Theory of Operation

The operation of the Flexostat (Figure 1) consists of a discrete dilution cycle with an adjustable period (typically 1 min). First, the current optical density of each chamber is measured. Then, for each chamber, a user-programmable Python module computes the dilution rate as a function of error between desired OD and measured OD. The dilution rates, which may be different for each chamber, are then sent to the embedded circuitry, which controls the media pump and valve system. As new media is added, it is mixed by magnetic stir bars diluting old media, cells, and waste products. When the media level rises to the mouth of the effluent tube, waste is forced out by positive pressure provided by an aeration input.

Specifications and Characterization

Biochemical reactions in cells are highly sensitive to experimental preparation, equipment variation, and environmental changes. Indeed, some systems even have multimodal

steady state distributions,(25) which are sensitive to initial conditions and require replicates to thoroughly probe. Furthermore, most experimental outcomes must be compared to those of a control experiment or between multiple variants. In a single chamber turbidostat, each replicate and control must be performed serially, which is time-consuming and susceptible to day to day variations in experimental conditions. For the Flexostat, we chose to multiplex eight culture chambers into the same media source and pump, which allows us to perform eight experiments in parallel, all of which are subjected to the same conditions. A single multiplexed pump has the additional advantage of lowering cost relative to buying eight pumps.

In the continuous culture literature, culture volumes can range from a few femtoliters(26) to industrial fermenters containing thousands of liters of media. The culture volume has logistical impacts that need to be carefully weighed. Yeast in rich media go through approximately 16 generations in a day if kept at maximal growth rates. Since one volume must be replaced for every generation and there are eight culture chambers, 16 generations/day/chamber × 1 volume/generation × 8 chambers = 128 volumes per day will be consumed, which quickly adds up for large culture volumes. Small volumes also have drawbacks. If sampling is required for characterization, enough effluent must be available for a single measurement. A typical flow cytometer for example requires at least 50 µL for measurement.(27) Sampling the 50 µL required for flow cytometry from a 1 mL vessel would require close to 5 min of collection time from effluent. We have found that a culture volume of 15 mL is a good balance, allowing for enough effluent to be collected over a sample period of 2 min while limiting the amount of media used to less than 2 L a day.

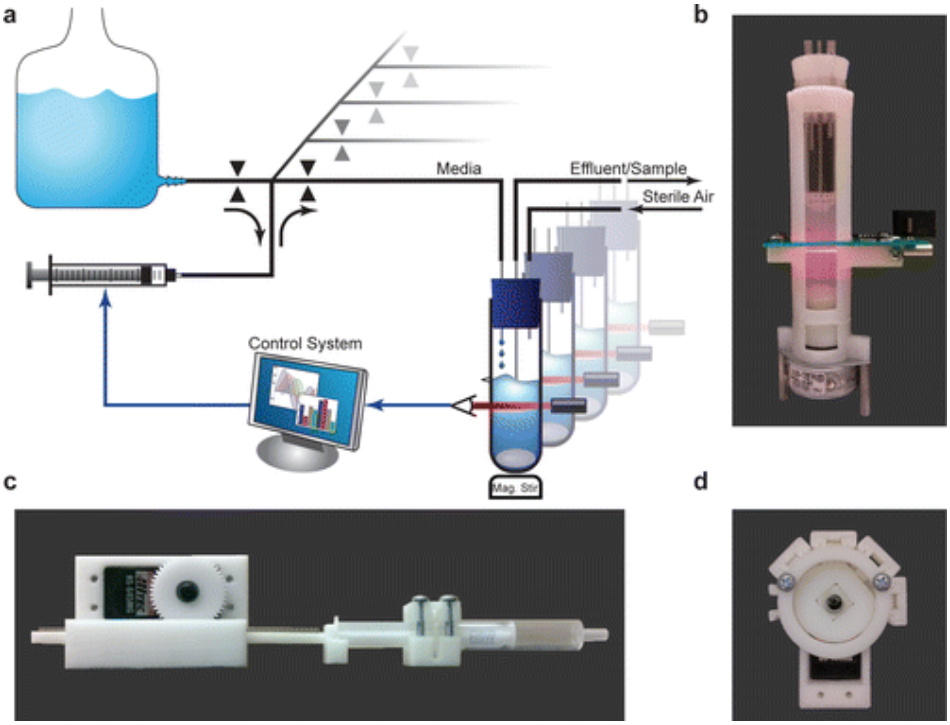


Figure 1. (a) Schematic of the Flexostat. Optical density (OD) is measured through the chamber wall and reported to a control system. A dilution rate is then calculated for each chamber, which is carried out by the pumping system. Valves select the flow source or destination while a single multiplexed syringe pump determines the volume and direction of flow. (b) A photograph of the turbidostat chamber with integrated OD measurement and stirring. (c) A 3D printed syringe pump. (d) A 3D printed four way normally closed pinch-valve with the upper right valve selected.

To measure error and validate the controller in the implemented system, we grew *S. cerevisiae* in duplicate at four different ODs and quantified the average, maximum, and 95th percentile tracking error (difference between set OD and measured OD) over a period of 22 h. Average tracking error was no more than 3.14×10^{-4} OD units between all eight chambers while the maximum error observed at any time point was $\pm 4.8 \times 10^{-3}$ OD units with 95% of the error inside of $\pm 2.3 \times 10^{-3}$OD units. Thus, accurate maintenance of a desired OD set point is achievable over experimentally relevant time periods.

Since the Flexostat is programmable through a Python application program interface (API), it can also be easily reprogrammed for different objectives (see Supporting Information section S1 for details). Dilution and feedback can be turned off allowing for the measurement of growth curves in the same device (Supporting Figure S3), which can be useful for determining batch growth rates and growth yields. It can also be programmed to follow arbitrary reference trajectories so long as they are physiologically feasible (Figure 3).

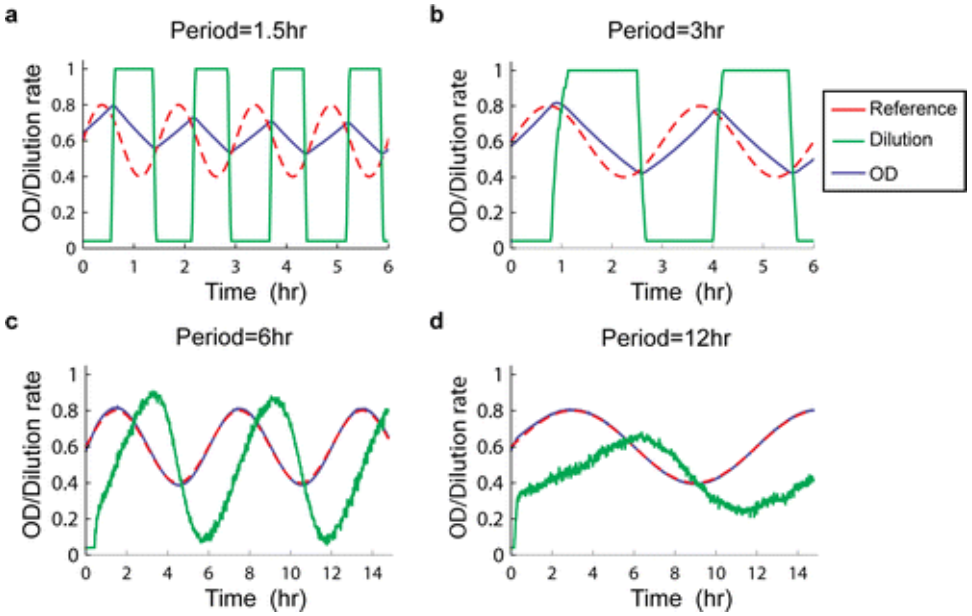


Figure 3. Experimental data collected from parallel experiments. *S. cerevisiae* was grown in synthetic complete media and made to track four sinusoidal reference trajectories (red). Measured OD values are shown in blue while the normalized dilution rate is shown in green. The maximum positive rate of change in OD is determined by the growth rate, while the maximum negative rate of change in OD is determined by the device's maximum dilution rate.

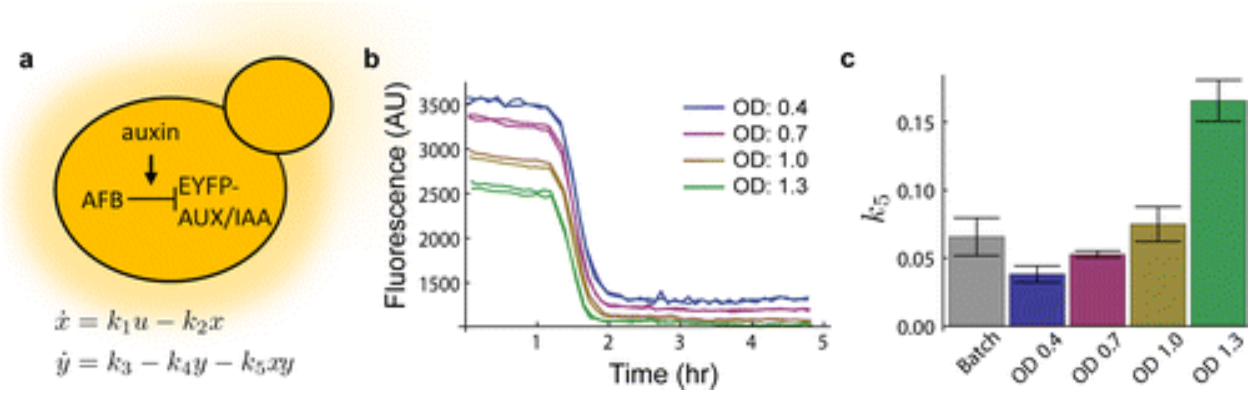


Figure 4. (a) Top: In the presence of auxin, AFB degrades the fusion protein EYFP-AUX/IAA. Bottom: A model from previous work(6) of the network shown. State x represents auxin bound to AFB with input u being the auxin concentration, parameter k_1 the association rate of auxin and AFB, and k_2 the natural degradation rate of AFB. State y represents the concentration of EYFP-AUX/IAA where k_3 is the production rate, k_4 is the natural degradation rate, and k_5 is the degradation rate in the presence of auxin bound to AFB. (b) YKL73 characterized in a turbidostat at four different ODs. At t = 1.25 h auxin was added to the culture chamber and media source to reach a concentration of 10 μM, and the response was measured by sampling effluent and reading mean fluorescence in a flow cytometer at a period of roughly 6 min. The cytometery data was gated to only include singlet (nonbudding) yeast in a healthy size range. An untreated time course is available as Supporting Information Figure S9 and shows no significant

change in fluorescence. (c) Model parameter k_5 was fit to the data collected in part b (See Havens et al.(6) Supplemental Data for detailed methods) and compared to previous work (gray).(6)

Gene Network Parameter Dependence on Cell Density

In previous work,(6) we studied the auxin plant hormone pathway cloned into *S. cerevisiae*. In our engineered yeast strain, as *in planta*, indole-3-acetic acid (auxin) promotes the interaction between a member of the F-box family of plant proteins (AFBs) and a member of the Aux/IAA family of plant proteins. AFBs target our engineered EYFP-Aux/IAA proteins to promote degradation via the proteasome (Figure 4a). In yeast transformed with a member of each family of AFB and EYFP-Aux/IAA proteins, strong YFP fluorescence can be measured through flow cytometery. With the addition of auxin, the EYFP-Aux/IAA fusion protein is degraded and within 2 h only background fluorescence is detectable. The exact degradation rates were characterized and shown to be dependent on the combination of AFBs and Aux/IAAs.

In previous work, each member of a combinatorial AFB Aux/IAA library was grown in synthetic complete medium (SC) under batch conditions at 30 °C in a shaker incubator overnight. The following day, the cultures were diluted and allowed to reenter exponential growth phase before induction at a cell count roughly equivalent to OD 0.4 in the Flexostat. Over the course of three or more hours, cells were sampled from batch and fluorescence was measured in a flow cytometer. These data were then fit to an ODE model (Figure 4a) where the parameter k_5 represents the degradation rate of EYFP-Aux/IAA in the presence of auxin bound to AFB.

We hypothesize that culture density has an effect on degradation rates either directly or indirectly through changes in nutrients, signaling molecules, and waste products that correspond

with cell density in a given media. To test our hypothesis, we selected a strain YKL73 from previous work that coexpresses AFB2 and EYFP-Aux/IAA6. YKL73 was grown in SC in our turbidostat inside of a 30 °C incubator at four different ODs: 0.4, 0.7, 1.0, and 1.3 in duplicate. Although we chose ODs ranging from 0.4 to 1.3, we did not observe a significant difference in growth rate between cultures, which is typical for growth at nonlimiting nutrient concentrations. To measure fluorescence in each chamber, 200 μL of effluent was sampled every 6 min and measured in the same flow cytometer as previous work (Figure 4b). After fitting the new degradation curves we found k_5 to vary from 0.036 ± 0.006, an average degrader, to 0.171 ± 0.015 (Figure 4c and Supporting Information Table S2), which is 35% faster than the fastest pair previously reported(6)from batch data.

Another apparent feature of the data are the initial AUX/IAA levels (quantity k_3/k_4 in the model). The initial expression levels also show an inverse relationship between OD and untreated expression of AUX/IAA. Figure 4c and Supporting Information Figures S10 and S11 illustrate the relationship between OD, initial expression, and degradation rates though the exact cause of the relationships is unknown.

In continuous culture, steady state levels of waste products and other exportable chemicals are proportional to cell density. Through the use of cytometer gating, we controlled for cell size and morphology (singlets vs budding cells). This leads us to hypothesize that either the rate k_5 and the ratio k_3/k_4 are highly sensitive to differences in growth rate (within the error margin of our measurements), or that the cells' internal state was affected by the density dependent effects in environment. Since k_5 is modeled as a constant, the density dependent effects, which could not have been measured in batch, highlight the utility of continuous culture in model validation and the need, in this case, for a higher order model to explain the data when

density cannot be controlled. Further study is required to determine the exact nature and source of the physiological changes that influence expression and degradation rates.

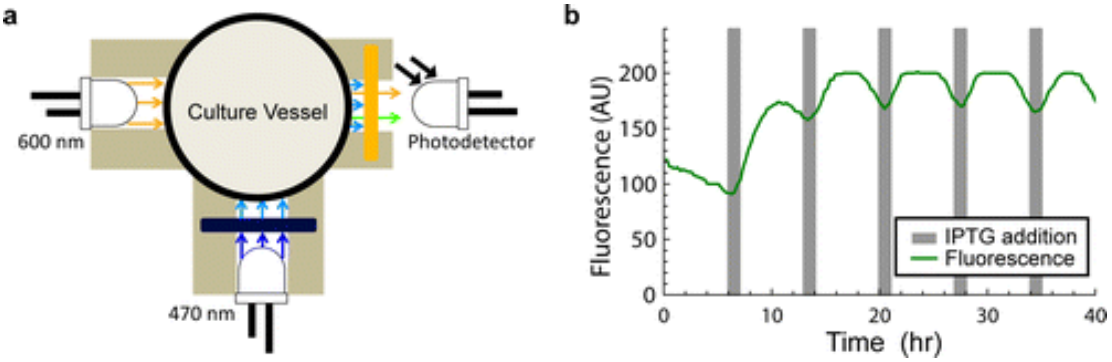


Figure 5. (a) Diagram representing the layout of the modified fluorostat culture chamber. The light sensor is time divison multiplexed between light from the absorbance source (600 nm) and fluorescence excited by the excitation source (470 nm). A filter set ensures that only light generated by fluorescence is detected. (b) An IPTG inducible GFP expressing strain of *E. coli* was grown at OD 0.6 in alternating M9 media with (gray area) and without IPTG.

Fluorescence Detecting Turbidostat

A key capability in synthetic and systems biology is the ability to read fluorescent reporters. In the previous section, we used a flow cytometer, which resulted in individual cell data but was labor intensive. An alternative method to collecting cytometry data is to obtain bulk fluorescence *in situ*, which can be collected without user intervention.

We modified the original chamber design to include a blue excitation light source at a right angle to the absorbance photo sensor (Figure 5a). The new design multiplexes the original light sensor in time to serve the dual purpose of measuring absorbance and fluorescence. For OD measurements, the chamber works as described in earlier sections. When fluorescence measurements are made, the 600 nm LED is turned off and the 470 nm LED excitation source is turned on. To prevent bleed over from the excitation source into the emissions wavelengths, we

added a band-pass optical filter (see Methods and Supporting Information section S4). Similarly, to prevent the excitation source from being detected by the photo sensor, a long-pass optical filter was added.

To test the fluorescence detection capability of our fluorescence detecting turbidostat (or Fluorostat for short), we measured a range of fluorescent cultures. Wild type *E. coli* (BL21) and BL21 containing a high expression super folder green fluorescent protein (sfGFP) plasmid were cultured for use in quantification. The cultures were then mixed at various ratios and measured in the Fluorostat, which showed a linear relationship between mix ratio and measured fluorescence (Supporting Information Figure S13).

Next, we tested the Fluorostat under more experimentally realistic conditions. We cultured an IPTG inducible sfGFP expressing strain of *E. coli* in the Fluorostat. At periods of 1 and 6 h, media containing 1 mM and 0 mM IPTG, respectively, were alternated and the resulting fluorescence signal was measured (Figure 5b). The results of this experiment demonstrate the ability of an LED based fluorimeter to read sfGFP reported outputs during the course of a turbidostat experiment. These results also highlight the adaptability of our design to work with evolving experimental needs.

While linear within its dynamic range and capable of measuring changes in expression levels, we discovered that only cultures bright enough to be visibly fluorescent under blue light transillumination were capable of being detected. The YFP expressing strain of yeast measured using cytometry in the previous section was too dimly fluorescent to measure in the Fluorostat. While its low sensitivity precludes the use of the Fluorostat in some experiments, we believe that through further refinement, the use of a more sensitive fluorometer, or an automatic cytometer

sampling system, all experiments requiring fluorescence measurement should be possible without manual sampling.

Conclusion

With the advent of 3D printing and other inexpensive custom fabrication processes, laboratories are now more capable of building their own equipment from open source designs than ever before. These community driven designs have the benefit of open design elements that allow laboratories to customize software and hardware for their own unique experiments rather than using designs optimized for industrial purposes. We presented a new 3D printable and customizable turbidostat for the next generation of "3D printed labs".

To demonstrate its utility, we have shown an example system with previously thought to be constant parameters that turned out to be dependent on cell density. We also demonstrated that our design can be adapted to accommodate fluorescence measurement, enabling the automatic collection of reporter expression data within the device.

References:

1. Canton, B., Labno, A., and Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices Nat. Biotechnol. 26, 787– 793
2. Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering Proc. Natl. Acad. Sci. U.S.A. 102, 12678– 12683
3. Olson, E., Hartsough, L., Landry, B., Shroff, R., and Tabor, J. (2014) Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals Nat. Methods 11, 449– 455
4. Ceroni, F., Furini, S., Stefan, A., Hochkoeppler, A., and Giordano, E. (2012) A synthetic post-transcriptional controller to explore the modular design of gene circuits ACS Synth. Biol. 1, 163– 171
5. Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K.,Glieberman, A. L., Monie, D. D., and Endy, D. (2009) Measuring the activity of BioBrick promoters using an*in vivo* reference standard J. Biol. Eng. 3, 4

6. Havens, K. A., Guseman, J. M., Jang, S. S., Pierre-Jerome, E., Bolten, N., Klavins, E., and Nemhauser, J. L.(2012) A synthetic approach reveals extensive tunability of auxin signaling Plant Physiol. 160, 135–142

7. Anderson, J. C., Voigt, C. A., and Arkin, A. P. (2007) Environmental signal integration by a modular AND gate Mol. Syst. Biol. 3, 133

8. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Mai, Q.-A., Christoffersen, M. J., Martin, L., Yu, A., Lam, C.,Rodriguez, C., and Bennett, G. 2013, Quantitative estimation of activity and quality for collections of functional genetic elements Nat. Methods 10, 347– 353

9. Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull,M., Keasling, J. D., and Arkin, A. P. 2013, Precise and reliable gene expression via standard transcription and translation initiation elements Nat. Methods 10, 354– 360

10. Monod, J. (1949) The growth of bacterial cultures Annu. Rev. Microbiol. 3, 371– 394

11. Brauer, M. J., Huttenhower, C., Airoldi, E. M., Rosenstein, R., Matese, J. C., Gresham, D., Boer, V. M.,Troyanskaya, O. G., and Botstein, D. (2008) Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast Mol. Biol. Cell 19, 352– 367

12. Valgepea, K., Adamberg, K., Seiman, A., and Vilu, R. (2013) *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins Mol. BioSyst. 9, 2344

13. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. (2010) Interdependence of cell growth and gene expression: Origins and consequences Science 330, 1099– 1102

14. Monod, J. (1950) La technique de culture continue, theorie et applications Ann. Inst. Pasteur 79, 390–410

15. Saldanha, A. J., Brauer, M. J., and Botstein, D. (2004) Nutritional homeostasis in batch and steady-state culture of yeast Mol. Biol. Cell 15, 4089– 4104

16. Tomson, K., Barber, J., and Vanatalu, K. (2006) Adaptastat—A new method for optimizing of bacterial growth conditions in continuous culture: Interactive substrate limitation based on dissolved oxygen measurement J. Microbiol. Methods 64, 380– 390

17. Toprak, E., Veres, A., Michel, J. B., Chait, R., Hartl, D. L., and Kishony, R. (2012) Evolutionary paths to antibiotic resistance under dynamically sustained drug selection Nat. Genet. 44, 101– 105

18. Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011) A system for the continuous directed evolution of biomolecules Nature 472, 499– 503

19. Lovitt, R. W. and Wimpenny, J. W. T. (1981) The gradostat: A bidirectional compound chemostat and its application in microbiological research Microbiology 127, 261– 268

20. Miller, A. W., Befort, C., Kerr, E. O., and Dunham, M. J. (2013) Design and use of multiplexed chemostat arrays J. Vis. Exp. e50262

21. Markx, G. H., Davey, C. L., and Kell, D. B. (1991) The permittistat: A novel type of turbidostat J. Gen. Microbiol. 137, 735– 743

22. Larsson, G., Enfors, S.-O., and Pham, H. (1990) The pH-auxostat as a tool for studying microbial dynamics in continuous fermentation Biotechnol. Bioeng. 36, 224– 232

23. Tappe, W., Tomaschewski, C., Rittershaus, S., and Groeneweg, J. (1996) Cultivation of nitrifying bacteria in the retentostat, a simple fermenter with internal biomass retention FEMS Microbiol. Ecol. 19, 47–52

24. Toprak, E., Veres, A., Yildiz, S., Pedraza, J. M., Chait, R., Paulsson, J., and Kishony, R. (2013) Building a morbidostat: An automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition Nat. Protoc. 8, 555– 567

25. Collins, J. J., Gardner, T. S., and Cantor, C. R. (2000) Construction of a genetic toggle switch in*Escherichia coli*. Nature 403, 339– 342

26. Wang, P., Robert, L., Pelletier, J., Dang, W. L., Taddei, F., Wright, A., and Jun, S. (2010) Robust growth of*Escherichia coli* Curr. Biol. 20, 1099– 1103

27. BD Accuri C6 Flow Cytometer Technical Specifications (2012) BD Biosciences, Piscataway, NJ.

28. Dorf, R. C. and Bishop, R. H. (2005) Modern Control Systems, 10th ed.; Prentice Hall, Upper Saddle River, NJ.

Appendix B: A high quality filtered list of mutations observed across 95 chemostat evolution experiments (Chapter 3).

Repeatedly observed mutations under glucose limitation

Population	Chr #	Position	Mut. Type	Sys name	Common name	Mutation	Proportion
G26	chrI	39261	coding-nonsynonymous	YAL056W	GBP2	M1I	0.204819277
G29	chrI	41731	coding-nonsynonymous	YAL056W	GBP2	E825*	0.965116279
G24	chrIV	1310265	coding-nonsynonymous	YDR420W	HKR1	K1333N	0.282051282
G31	chrIV	1308659	coding-nonsynonymous	YDR420W	HKR1	F798Y	0.204081633
G15	chrV	1645	coding-nonsynonymous	YEL077C		N818T	0.454545455
G15	chrV	1769	coding-nonsynonymous	YEL077C		K777*	0.413793103

G26	chrV	1717	coding-nonsynonymous	YEL077C		N794I	0.523809524
G18	chrV	479831	coding-nonsynonymous	YER155C	BEM2	Y1006*	0.197916667
G18	chrV	479940	coding-nonsynonymous	YER155C	BEM2	S970*	0.148648649
G19	chrV	477648	coding-nonsynonymous	YER155C	BEM2	I1734indel	0.406593407
G19	chrV	480767	coding-nonsynonymous	YER155C	BEM2	W694*	0.146666667
G20	chrV	481748	coding-nonsynonymous	YER155C	BEM2	Q367indel	0.786885246
G22	chrV	480741	coding-nonsynonymous	YER155C	BEM2	S703*	0.955555556
G23	chrV	476603	coding-nonsynonymous	YER155C	BEM2	Y2082*	0.633333333
G29	chrV	479338	coding-nonsynonymous	YER155C	BEM2	A1171indel	0.382352941
G29	chrV	481111	coding-nonsynonymous	YER155C	BEM2	E580*	0.225225225
G2	chrVI	70130	coding-nonsynonymous	YFL033C	RIM15	L1433*	0.485714286
G6	chrVI	71985	coding-nonsynonymous	YFL033C	RIM15	L815indel	0.4
G1	chrIX	69887	coding-nonsynonymous	YIL147C	SLN1	M1189I	0.166666667
G25	chrIX	70005	coding-nonsynonymous	YIL147C	SLN1	V1150G	0.618181818
G26	chrIX	71790	coding-nonsynonymous	YIL147C	SLN1	A555V	0.402061856
G4	chrIX	69868	coding-nonsynonymous	YIL147C	SLN1	P1196T	0.610619469
G10	chrX	236338	coding-nonsynonymous	YJL101C	GSH1	G7S	0.1875
G11	chrX	234838	coding-nonsynonymous	YJL101C	GSH1	Y507N	0.224137931
G17	chrX	235255	coding-nonsynonymous	YJL101C	GSH1	F368I	0.47761194
G18	chrX	236005	coding-nonsynonymous	YJL101C	GSH1	V118F	0.885714286
G19	chrX	235587	coding-nonsynonymous	YJL101C	GSH1	S257F	0.984375
G20	chrX	235571	coding-nonsynonymous	YJL101C	GSH1	M262I	0.989795918
G22	chrX	235723	coding-nonsynonymous	YJL101C	GSH1	V212F	0.267857143
G26	chrX	235519	coding-nonsynonymous	YJL101C	GSH1	R280G	0.2
G29	chrX	234451	coding-nonsynonymous	YJL101C	GSH1	Y636D	0.925233645
G31	chrX	235539	coding-nonsynonymous	YJL101C	GSH1	A273E	0.355932203
G31	chrX	236005	coding-nonsynonymous	YJL101C	GSH1	V118F	0.238095238
G5	chrX	235994	coding-nonsynonymous	YJL101C	GSH1	N121K	0.527777778
G19	chrXII	406479	coding-nonsynonymous	YLR131C	ACE2	S115indel	0.971428571
G20	chrXII	406838	5'-upstream	YLR131C	ACE2	NA	0.975806452
G22	chrXII	406111	coding-nonsynonymous	YLR131C	ACE2	Q238*	0.888888889
G23	chrXII	406820	coding-nonsynonymous	YLR131C	ACE2	M1I	0.87654321
G29	chrXII	404817	coding-nonsynonymous	YLR131C	ACE2	R669indel	0.942857143
G18	chrXIV	172379	5'-upstream	YNL252C	MRPL17	NA	0.261904762
G32	chrXIV	172382	5'-upstream	YNL252C	MRPL17	NA	0.314285714
G20	chrXIV	704334	coding-synonymous	YNR044W	AGA1	T212T	0.304347826
G20	chrXIV	704355	coding-synonymous	YNR044W	AGA1	T219T	0.274509804
G32	chrXV	216940	5'-upstream	YOL059W	GPD2	NA	0.5
G32	chrXV	216956	5'-upstream	YOL059W	GPD2	NA	0.447368421
G10	chrXV	411326	coding-nonsynonymous	YOR043W	WHI2	E153*	0.195652174
G12	chrXV	411594	coding-nonsynonymous	YOR043W	WHI2	S242*	0.336538462
G14	chrXV	412193	coding-nonsynonymous	YOR043W	WHI2	E442*	0.242857143
G27	chrXV	411452	coding-nonsynonymous	YOR043W	WHI2	R195*	0.5

G27	chrXV	411557	coding-nonsynonymous	YOR043W	WHI2	Q230indel	0.264705882
G27	chrXV	411563	coding-nonsynonymous	YOR043W	WHI2	Q232K	0.271428571
G3	chrXV	411452	coding-nonsynonymous	YOR043W	WHI2	R195*	0.173913043
G27	chrXV	1036390	coding-nonsynonymous	YOR372C	NDD1	N28indel	0.43956044
G31	chrXV	1036451	coding-nonsynonymous	YOR372C	NDD1	Y7*	0.278481013

Repeatedly observed mutations under phosphate-limitation

Population	Chr #	Position	Mut. Type	Sys name	Common name	Mutation	Proportion
P16	chrII	798420	coding-nonsynonymous	YBR296C	PHO89	S35A	0.206349206
P20	chrII	798420	coding-nonsynonymous	YBR296C	PHO89	S35A	0.23943662
P26	chrII	798420	coding-nonsynonymous	YBR296C	PHO89	S35A	0.214285714
P7	chrII	798080	coding-nonsynonymous	YBR296C	PHO89	S148*	0.251655629
P10	chrIV	427414	coding-synonymous	YDL014W	NOP1	G17G	0.333333333
P24	chrIV	427593	coding-nonsynonymous	YDL014W	NOP1	A77G	0.322580645
P20	chrIV	378121	coding-nonsynonymous	YDL042C	SIR2	T109P	0.203703704
P26	chrIV	378422	coding-nonsynonymous	YDL042C	SIR2	Y8*	0.188679245
P19	chrIV	205771	coding-synonymous	YDL140C	RPO21	S1597S	0.4
P5	chrIV	209881	coding-synonymous	YDL140C	RPO21	V227V	0.14
P13	chrIV	918933	coding-nonsynonymous	YDR227W	SIR4	E455*	0.549019608
P8	chrIV	920313	coding-nonsynonymous	YDR227W	SIR4	Q915*	0.405405405
P15	chrV	299655	coding-nonsynonymous	YER070W	RNR1	K236E	0.864864865
P26	chrV	299397	coding-nonsynonymous	YER070W	RNR1	T150P	0.140495868
P26	chrV	300466	coding-nonsynonymous	YER070W	RNR1	A506D	0.432098765
P1	chrV	506640	coding-nonsynonymous	YER164W	CHD1	W417indel	0.595833333
P20	chrV	506319	coding-nonsynonymous	YER164W	CHD1	K310E	0.255319149
P20	chrV	506394	coding-nonsynonymous	YER164W	CHD1	V335indel	0.217948718
P30	chrV	507507	coding-nonsynonymous	YER164W	CHD1	R706*	0.774647887
P5	chrV	505587	coding-nonsynonymous	YER164W	CHD1	E66*	0.157407407
P7	chrV	508903	coding-nonsynonymous	YER164W	CHD1	W1171*	0.199074074
P21	chrVII	583069	coding-nonsynonymous	YGR044C	RME1	R275K	0.880733945
P23	chrVII	583193	coding-nonsynonymous	YGR044C	RME1	H234Y	0.226666667
P29	chrVII	583519	coding-nonsynonymous	YGR044C	RME1	Q125indel	0.357142857
P3	chrVII	583105	coding-nonsynonymous	YGR044C	RME1	C263S	0.210526316
P4	chrVII	583036	coding-nonsynonymous	YGR044C	RME1	S286*	0.897435897
P5	chrVII	583892	coding-nonsynonymous	YGR044C	RME1	M1L	0.420454545
P20	chrVIII	512789	coding-nonsynonymous	YHR206W	SKN7	N20Y	0.277777778
P22	chrVIII	512783	coding-nonsynonymous	YHR206W	SKN7	N18Y	0.253968254
P32	chrVIII	512783	coding-nonsynonymous	YHR206W	SKN7	N18Y	0.279069767
P14	chrXI	642185	coding-nonsynonymous	YKR101W	SIR1	L549*	0.182795699
P29	chrXI	641265	coding-nonsynonymous	YKR101W	SIR1	N242indel	0.2
P9	chrXI	641180	coding-nonsynonymous	YKR101W	SIR1	S214*	0.418439716
P14	chrXII	1019664	coding-nonsynonymous	YLR442C	SIR3	R863I	0.212765957

P14	chrXII	1022219	coding-nonsynonymous	YLR442C	SIR3	W11*	0.348837209
P8	chrXII	1020295	coding-nonsynonymous	YLR442C	SIR3	V653F	0.313653137
P1	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.920792079
P10	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.59602649
P11	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.516129032
P12	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.988505747
P13	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.518072289
P15	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.877358491
P16	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.277777778
P17	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	1
P19	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.275862069
P2	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.841772152
P20	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.4
P21	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.982142857
P24	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.2
P26	chrXIII	25024	coding-synonymous	YML123C	PHO84	L259L	0.5125
P26	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.93902439
P27	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.764044944
P29	chrXIII	25024	coding-synonymous	YML123C	PHO84	L259L	0.4375
P29	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.507936508
P30	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.862068966
P4	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.908
P5	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.539130435
P7	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.366834171
P8	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.80075188
P9	chrXIII	25025	coding-nonsynonymous	YML123C	PHO84	L259P	0.564102564
P20	chrXIV	602291	coding-synonymous	YNL017C		P62P	0.180327869
P20	chrXIV	602296	coding-nonsynonymous	YNL017C		S61P	0.171875
P20	chrXIV	602310	coding-nonsynonymous	YNL017C		I56T	0.179104478
P20	chrXIV	600871	coding-nonsynonymous	YNL018C		T302A	0.206349206
P20	chrXIV	600876	coding-nonsynonymous	YNL018C		V300G	0.164556962
P10	chrXV	442477	coding-nonsynonymous	YOR061W	CKA2	L315*	0.239726027
P14	chrXV	441777	coding-nonsynonymous	YOR061W	CKA2	K82*	0.343137255
P23	chrXV	441753	coding-nonsynonymous	YOR061W	CKA2	Q74*	0.280701754
P25	chrXV	442216	coding-nonsynonymous	YOR061W	CKA2	S228*	0.666666667
P29	chrXV	441663	coding-nonsynonymous	YOR061W	CKA2	W44indel	0.369565217
P30	chrXVI	607448	5'-upstream	YPR022C	SDD4	NA	0.326530612
P32	chrXVI	607421	5'-upstream	YPR022C	SDD4	NA	0.428571429

Repeatedly observed mutations under sulfate-limitation

Population	Chr #	Position	Mut. Type	Sys name	Common name	Mutation	Proportion
S13	chrII	789984	coding-nonsynonymous	YBR294W	SUL1	N250K	0.812206573

S17	chrII	790023	coding-nonsynonymous	YBR294W	SUL1	N263K	0.6996337
S19	chrII	789984	coding-nonsynonymous	YBR294W	SUL1	N250K	0.367346939
S2	chrII	789984	coding-nonsynonymous	YBR294W	SUL1	N250K	1
S20	chrII	790023	coding-nonsynonymous	YBR294W	SUL1	N263K	1
S21	chrII	790021	coding-nonsynonymous	YBR294W	SUL1	N263H	0.982935154
S23	chrII	789984	coding-nonsynonymous	YBR294W	SUL1	N250K	0.685446009
S24	chrII	790021	coding-nonsynonymous	YBR294W	SUL1	N263H	0.8515625
S25	chrII	789984	coding-nonsynonymous	YBR294W	SUL1	N250K	0.913294798
S26	chrII	789984	coding-nonsynonymous	YBR294W	SUL1	N250K	0.70625
S26	chrII	790023	coding-nonsynonymous	YBR294W	SUL1	N263K	0.316770186
S27	chrII	790023	coding-nonsynonymous	YBR294W	SUL1	N263K	0.976415094
S29	chrII	789984	coding-nonsynonymous	YBR294W	SUL1	N250K	0.730434783
S6	chrII	789984	coding-nonsynonymous	YBR294W	SUL1	N250K	0.906810036
S19	chrIV	759578	coding-synonymous	YDR150W	NUM1	V1317V	0.451612903
S20	chrIV	759566	coding-synonymous	YDR150W	NUM1	L1313L	0.323529412
S20	chrIV	759597	coding-nonsynonymous	YDR150W	NUM1	N1324Y	0.357142857
S29	chrIV	759506	coding-nonsynonymous	YDR150W	NUM1	H1293Q	0.363636364
S30	chrIV	757544	coding-synonymous	YDR150W	NUM1	H639H	0.391304348
S30	chrIV	758882	coding-nonsynonymous	YDR150W	NUM1	E1085D	0.391304348
S26	chrIV	1259709	coding-nonsynonymous	YDR392W	SPT3	*338Y	0.164179104
S26	chrIV	1259709	5'-upstream	YDR393W	SPT3	NA	0.164179104
S1	chrVII	378682	coding-nonsynonymous	YGL066W	SGF73	T358indel	0.205479452
S11	chrVII	378686	coding-nonsynonymous	YGL066W	SGF73	Q360*	0.657142857
S12	chrVII	378345	coding-nonsynonymous	YGL066W	SGF73	S246*	0.333333333
S12	chrVII	378764	coding-nonsynonymous	YGL066W	SGF73	K386*	0.322033898
S14	chrVII	378409	coding-nonsynonymous	YGL066W	SGF73	Y267*	0.571428571
S16	chrVII	378764	coding-nonsynonymous	YGL066W	SGF73	K386*	0.884615385
S17	chrVII	377746	coding-nonsynonymous	YGL066W	SGF73	Y46*	0.157894737
S18	chrVII	378437	coding-nonsynonymous	YGL066W	SGF73	E277*	0.311827957
S2	chrVII	377779	coding-nonsynonymous	YGL066W	SGF73	Y57*	0.328358209
S22	chrVII	378242	coding-nonsynonymous	YGL066W	SGF73	K212*	0.313953488
S30	chrVII	378334	coding-nonsynonymous	YGL066W	SGF73	Y242*	0.16
S30	chrVII	378488	coding-nonsynonymous	YGL066W	SGF73	E294*	0.735632184
S4	chrVII	378173	coding-nonsynonymous	YGL066W	SGF73	N189indel	0.728813559
S7	chrVII	378368	coding-nonsynonymous	YGL066W	SGF73	M254indel	0.661016949
S8	chrVII	378650	coding-nonsynonymous	YGL066W	SGF73	G348*	0.333333333
S9	chrVII	378437	coding-nonsynonymous	YGL066W	SGF73	E277*	0.965909091
S2	chrIX	23872	coding-synonymous	YIL169C		T745T	0.277777778
S30	chrIX	25719	coding-nonsynonymous	YIL169C		T130S	0.291666667
S9	chrIX	25802	coding-nonsynonymous	YIL169C		V102D	0.4
S16	chrX	477659	coding-synonymous	YJR027W		V1633V	0.461538462
S16	chrX	477674	coding-synonymous	YJR027W		S1638S	0.4375
S16	chrX	477677	coding-synonymous	YJR027W		T1639T	0.431578947

S24	chrXIII	908484	coding-nonsynonymous	YMR317W		L374*	0.48
S30	chrXIII	908275	coding-synonymous	YMR317W		T304T	0.25
S31	chrXIII	908232	coding-nonsynonymous	YMR317W		L290*	0.454545455
S20	chrXV	1060708	coding-synonymous	YOR383C	FIT3	H116H	0.275
S31	chrXV	1060732	coding-synonymous	YOR383C	FIT3	T108T	0.2
S26	chrXVI	69359	5'-upstream	YPL254W	HFI1	NA	0.23255814
S3	chrXVI	70542	coding-nonsynonymous	YPL254W	HFI1	L353*	0.139240506

BIBLIOGRAPHY

Aminetzach, Y. T., J. M. Macpherson and D. A. Petrov (2005). "Pesticide resistance via transposition-mediated adaptive gene truncation in Drosophila." Science **309**(5735): 764-767.

Arlow, T., K. Scott, A. Wagenseller and A. Gammie (2013). "Proteasome inhibition rescues clinically significant unstable variants of the mismatch repair protein Msh2." Proc Natl Acad Sci U S A **110**(1): 246-251.

Baker, S. P. and P. A. Grant (2007). "The SAGA continues: expanding the cellular role of a transcriptional co-activator complex." Oncogene **26**(37): 5329-5340.

Balagadde, F. K., L. You, C. L. Hansen, F. H. Arnold and S. R. Quake (2005). "Long-term monitoring of bacteria undergoing programmed population control in a microchemostat." Science **309**(5731): 137-140.

Barrick, J. E. and R. E. Lenski (2009). "Genome-wide mutational diversity in an evolving population of Escherichia coli." Cold Spring Harb Symp Quant Biol **74**: 119-129.

Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski and J. F. Kim (2009). "Genome evolution and adaptation in a long-term experiment with Escherichia coli." Nature **461**(7268): 1243-1247.

Barroso-Batista, J., A. Sousa, M. Lourenco, M. L. Bergman, D. Sobral, J. Demengeot, K. B. Xavier and I. Gordo (2014). "The first steps of adaptation of Escherichia coli to the gut are dominated by soft sweeps." PLoS Genet **10**(3): e1004182.

Boer, V. M., J. H. de Winde, J. T. Pronk and M. D. Piper (2003). "The genome-wide transcriptional responses of Saccharomyces cerevisiae grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur." J Biol Chem **278**(5): 3265-3274.

Bolger, A. M., M. Lohse and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics **30**(15): 2114-2120.

Brauer, M. J., A. J. Saldanha, K. Dolinski and D. Botstein (2005). "Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures." Mol Biol Cell **16**(5): 2503-2517.

Brown, C. J., K. M. Todd and R. F. Rosenzweig (1998). "Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment." Mol Biol Evol **15**(8): 931-942.

Castrillo, J. I., L. A. Zeef, D. C. Hoyle, N. Zhang, A. Hayes, D. C. Gardner, M. J. Cornell, J. Petty, L. Hakes, L. Wardleworth, B. Rash, M. Brown, W. B. Dunn, D. Broadhurst, K.

O'Donoghue, S. S. Hester, T. P. Dunkley, S. R. Hart, N. Swainston, P. Li, S. J. Gaskell, N. W. Paton, K. S. Lilley, D. B. Kell and S. G. Oliver (2007). "Growth control of the eukaryote cell: a systems biology study in yeast." J Biol **6**(2): 4.

Cherest, H., J. C. Davidian, D. Thomas, V. Benes, W. Ansorge and Y. Surdin-Kerjan (1997). "Molecular characterization of two high affinity sulfate transporters in Saccharomyces cerevisiae." Genetics **145**(3): 627-635.

Chiotti, K. E., D. J. Kvitek, K. H. Schmidt, G. Koniges, K. Schwartz, E. A. Donckels, F. Rosenzweig and G. Sherlock (2014). "The Valley-of-Death: reciprocal sign epistasis constrains adaptive trajectories in a constant, nutrient limiting environment." Genomics **104**(6 Pt A): 431-437.

Daran-Lapujade, P., J. M. Daran, A. J. van Maris, J. H. de Winde and J. T. Pronk (2009). "Chemostat-based micro-array analysis in baker's yeast." Adv Microb Physiol **54**: 257-311.

de Kok, S., J. F. Nijkamp, B. Oud, F. C. Roque, D. de Ridder, J. M. Daran, J. T. Pronk and A. J. van Maris (2012). "Laboratory evolution of new lactate transporter genes in a jen1Delta mutant of Saccharomyces cerevisiae and their identification as ADY2 alleles by whole-genome resequencing and transcriptome analysis." FEMS Yeast Res.

de la Chapelle, A. (2005). "The incidence of Lynch syndrome." Fam Cancer **4**(3): 233-237.
Diderich, J. A., M. Schepper, P. van Hoek, M. A. Luttik, J. P. van Dijken, J. T. Pronk, P. Klaassen, H. F. Boelens, M. J. de Mattos, K. van Dam and A. L. Kruckeberg (1999). "Glucose uptake kinetics and transcription of HXT genes in chemostat cultures of Saccharomyces cerevisiae." J Biol Chem **274**(22): 15350-15359.

Dorsey, M., C. Peterson, K. Bray and C. E. Paquin (1992). "Spontaneous amplification of the ADH4 gene in Saccharomyces cerevisiae." Genetics **132**(4): 943-950.
Dunham, M. J. (2010). "Experimental evolution in yeast: a practical guide." Methods Enzymol **470**: 487-507.

Ferenci, T. (2008). "Bacterial physiology, regulation and mutational adaptation in a chemostat environment." Adv Microb Physiol **53**: 169-229.

Gammie, A. E., N. Erdeniz, J. Beaver, B. Devlin, A. Nanji and M. D. Rose (2007). "Functional characterization of pathogenic human MSH2 missense mutations in Saccharomyces cerevisiae." Genetics **177**(2): 707-721.

Gilbert, L. A., M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman and L. S. Qi (2013). "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes." Cell **154**(2): 442-451.

Gresham, D., M. M. Desai, C. M. Tucker, H. T. Jenq, D. A. Pai, A. Ward, C. G. DeSevo, D. Botstein and M. J. Dunham (2008). "The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast." PLoS Genet **4**(12): e1000303.

Gresham, D., R. Usaite, S. M. Germann, M. Lisby, D. Botstein and B. Regenberg (2010). "Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAP1 locus." Proc Natl Acad Sci U S A **107**(43): 18551-18556.

Groisman, A., C. Lobo, H. Cho, J. K. Campbell, Y. S. Dufour, A. M. Stevens and A. Levchenko (2005). "A microfluidic chemostat for experiments with bacterial and yeast cells." Nat Methods **2**(9): 685-689.

Helling, R. B., C. N. Vargas and J. Adams (1987). "Evolution of Escherichia coli during growth in a constant environment." Genetics **116**(3): 349-358.

Ho, C. H., L. Magtanong, S. L. Barker, D. Gresham, S. Nishimura, P. Natarajan, J. L. Koh, J. Porter, C. A. Gray, R. J. Andersen, G. Giaever, C. Nislow, B. Andrews, D. Botstein, T. R. Graham, M. Yoshida and C. Boone (2009). "A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds." Nat Biotechnol **27**(4): 369-377.

Hong, J. and D. Gresham (2014). "Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments." PLoS Genet **10**(1): e1004041.
Hoskisson, P. A. and G. Hobbs (2005). "Continuous culture--making a comeback?" Microbiology **151**(Pt 10): 3153-3159.

Hughes, T. R., C. J. Roberts, H. Dai, A. R. Jones, M. R. Meyer, D. Slade, J. Burchard, S. Dow, T. R. Ward, M. J. Kidd, S. H. Friend and M. J. Marton (2000). "Widespread aneuploidy revealed by DNA microarray expression profiling." Nat Genet **25**(3): 333-337.

Jones, G. M., J. Stalker, S. Humphray, A. West, T. Cox, J. Rogers, I. Dunham and G. Prelich (2008). "A systematic library for comprehensive overexpression screens in Saccharomyces cerevisiae." Nat Methods **5**(3): 239-241.

Kao, K. C. and G. Sherlock (2008). "Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae." Nat Genet **40**(12): 1499-1504.

Kariola, R., R. Otway, K. E. Lonnqvist, T. E. Raevaara, F. Macrae, Y. J. Vos, M. Kohonen-Corish, R. M. Hofstra and M. Nystrom-Lahti (2003). "Two mismatch repair gene mutations found in a colon cancer patient--which one is pathogenic?" Hum Genet **112**(2): 105-109.

Klein, T., K. Schneider and E. Heinzle (2013). "A system of miniaturized stirred bioreactors for parallel continuous cultivation of yeast with online measurement of dissolved oxygen and off-gas." Biotechnol Bioeng **110**(2): 535-542.

Kubitschek, H. E. and H. E. Bendigkeit (1964). "MUTATION IN CONTINUOUS CULTURES. I. DEPENDENCE OF MUTATIONAL RESPONSE UPON GROWTH-LIMITING FACTORS." Mutat Res **106**: 113-120.

Kunkel, T. A. and D. A. Erie (2005). "DNA mismatch repair." Annu Rev Biochem **74**: 681-710.

Kvitek, D. J. and G. Sherlock (2011). "Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape." PLoS Genet **7**(4): e1002056.

Kvitek, D. J. and G. Sherlock (2013). "Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment." PLoS Genet **9**(11): e1003972.

Laan, L., J. H. Koschwanez and A. W. Murray (2015). "Evolutionary adaptation after crippling cell polarization follows reproducible trajectories." Elife **4**.

Lang, G. I. and A. W. Murray (2008). "Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae." Genetics **178**(1): 67-82.

Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein and M. M. Desai (2013). "Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations." Nature **500**(7464): 571-574.

Levy, S., M. Kafri, M. Carmi and N. Barkai (2011). "The Competitive Advantage of a Dual-Transporter System." Science **334**(6061): 1408-1412.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Lipton, L. R., V. Johnson, C. Cummings, S. Fisher, P. Risby, A. T. Eftekhar Sadat, T. Cranston, L. Izatt, P. Sasieni, S. V. Hodgson, H. J. Thomas and I. P. Tomlinson (2004). "Refining the Amsterdam Criteria and Bethesda Guidelines: testing algorithms for the prediction of mismatch repair mutation status in the familial cancer clinic." J Clin Oncol **22**(24): 4934-4943.

Liu, X., X. Jian and E. Boerwinkle (2011). "dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions." Hum Mutat **32**(8): 894-899.

Luria, S. E. and M. Delbruck (1943). "Mutations of Bacteria from Virus Sensitivity to Virus Resistance." Genetics **28**(6): 491-511.

Lynch, H. T., S. J. Lanspa, B. M. Boman, T. Smyrk, P. Watson, J. F. Lynch, P. M. Lynch, G. Cristofaro, P. Bufo, A. V. Tauro and et al. (1988). "Hereditary nonpolyposis colorectal cancer--Lynch syndromes I and II." Gastroenterol Clin North Am **17**(4): 679-712.

Lynch, H. T. and J. F. Lynch (1993). "The Lynch syndromes." Curr Opin Oncol **5**(4): 687-696.
Mecklin, J. P. and H. J. Jarvinen (1991). "Tumor spectrum in cancer family syndrome (hereditary nonpolyposis colorectal cancer)." Cancer **68**(5): 1109-1112.

Miller, A. W., C. Befort, E. O. Kerr and M. J. Dunham (2013). "Design and use of multiplexed chemostat arrays." J Vis Exp(72): e50262.

Monod, J. (1950). "La technique de culture continue. Théorie et applications." Ann Inst Pasteur. **79**: 390–410.

Nanchen, A., A. Schicker and U. Sauer (2006). "Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of Escherichia coli." Appl Environ Microbiol **72**(2): 1164-1172.

Novick, A. and L. Szilard (1950). "Description of the chemostat." Science **112**(2920): 715-716.

Novick, A. and L. Szilard (1950). "Experiments with the Chemostat on spontaneous mutations of bacteria." Proc Natl Acad Sci U S A **36**(12): 708-719.

Omasits, U., C. H. Ahrens, S. Muller and B. Wollscheid (2014). "Protter: interactive protein feature visualization and integration with experimental proteomic data." Bioinformatics **30**(6): 884-886.

Paquin, C. E. and J. Adams (1983). "Relative fitness can decrease in evolving asexual populations of S. cerevisiae." Nature **306**(5941): 368-370.

Payen, C., S. C. Di Rienzi, G. T. Ong, J. L. Pogachar, J. C. Sanchez, A. B. Sunshine, M. K. Raghuraman, B. J. Brewer and M. J. Dunham (2014). "The Dynamics of Diverse Segmental Amplifications in Populations of Saccharomyces cerevisiae Adapting to Strong Selection." G3: Genes|Genomes|Genetics **4**(3): 399-409.

Payen, C., S. C. Di Rienzi, G. T. Ong, J. L. Pogachar, J. C. Sanchez, A. B. Sunshine, M. K. Raghuraman, B. J. Brewer and M. J. Dunham (2014). "The dynamics of diverse segmental amplifications in populations of Saccharomyces cerevisiae adapting to strong selection." G3 (Bethesda) **4**(3): 399-409.

Payen, C., A. B. Sunshine, G. T. Ong, J. L. Pogachar, W. Zhao and M. J. Dunham (2015). "Empirical determinants of adaptive mutations in yeast experimental evolution." bioRxiv.

Peltomaki, P. and H. Vasen (2004). "Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database." Dis Markers **20**(4-5): 269-276.

Perlstein, E. O., D. M. Ruderfer, D. C. Roberts, S. L. Schreiber and L. Kruglyak (2007). "Genetic basis of individual differences in the response to small-molecule drugs in yeast." Nat Genet **39**(4): 496-502.

Perry, G. H., N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee and A. C. Stone (2007). "Diet and the evolution of human amylase gene copy number variation." Nat Genet **39**(10): 1256-1260.

Qian, Z., H. Huang, J. Y. Hong, C. L. Burck, S. D. Johnston, J. Berman, A. Carol and S. W. Liebman (1998). "Yeast Ty1 retrotransposition is stimulated by a synergistic interaction between mutations in chromatin assembly factor I and histone regulatory proteins." Mol Cell Biol **18**(8): 4783-4792.

Regenberg, B., T. Grotkjaer, O. Winther, A. Fausboll, M. Akesson, C. Bro, L. K. Hansen, S. Brunak and J. Nielsen (2006). "Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae." Genome Biol **7**(11): R107.

Saldanha, A. J., M. J. Brauer and D. Botstein (2004). "Nutritional Homeostasis in Batch and Steady-State Culture of Yeast." Molecular Biology of the Cell **15**(9): 4089-4104.

Siegel, R., D. Naishadham and A. Jemal (2012). "Cancer statistics, 2012." CA Cancer J Clin **62**(1): 10-29.

Sjursen, W., B. I. Haukanes, E. M. Grindedal, H. Aarset, A. Stormorken, L. F. Engebretsen, C. Jonsrud, I. Bjornevoll, P. A. Andresen, S. Ariansen, L. A. Lavik, B. Gilde, I. M. Bowitz-Lothe, L. Maehle and P. Moller (2010). "Current clinical criteria for Lynch syndrome are not sensitive enough to identify MSH6 mutation carriers." J Med Genet **47**(9): 579-585.

Steiner, C. C., J. N. Weber and H. E. Hoekstra (2007). "Adaptive Variation in Beach Mice Produced by Two Interacting Pigmentation Genes." PLoS Biol **5**(9): e219.

Sunshine, A. B., C. Payen, G. T. Ong, I. Liachko, K. M. Tan and M. J. Dunham (2015). "The fitness consequences of aneuploidy are driven by condition-dependent gene effects." PLoS Biol **13**(5): e1002155.

Swallow, J. G., P. Koteja, P. A. Carter and T. Garland (1999). "Artificial selection for increased wheel-running activity in house mice results in decreased body mass at maturity." J Exp Biol **202**(Pt 18): 2513-2520.

Szamecz, B., G. Boross, D. Kalapis, K. Kovacs, G. Fekete, Z. Farkas, V. Lazar, M. Hrtyan, P. Kemmeren, M. J. Groot Koerkamp, E. Rutkai, F. C. Holstege, B. Papp and C. Pal (2014). "The genomic landscape of compensatory evolution." PLoS Biol **12**(8): e1001935.

Takahashi, C. N., A. W. Miller, F. Ekness, M. J. Dunham and E. Klavins (2015). "A low cost, customizable turbidostat for use in synthetic circuit characterization." ACS Synth Biol **4**(1): 32-38.

Takahashi, M., H. Shimodaira, C. Andreutti-Zaugg, R. Iggo, R. D. Kolodner and C. Ishioka (2007). "Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays." Cancer Res **67**(10): 4595-4604.

Tenaillon, O., A. Rodriguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long and B. S. Gaut (2012). "The molecular diversity of adaptive convergence." Science **335**(6067): 457-461.

Tennessen, J. A., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, S. Gravel, S. McGee, R. Do, X. Liu, G. Jun, H. M. Kang, D. Jordan, S. M. Leal, S. Gabriel, M. J. Rieder, G. Abecasis, D. Altshuler, D. A. Nickerson, E. Boerwinkle, S. Sunyaev, C. D. Bustamante, M. J. Bamshad and J. M. Akey (2012). "Evolution and functional impact of rare coding variation from deep sequencing of human exomes." Science **337**(6090): 64-69.

Teotonio, H., I. M. Chelo, M. Bradic, M. R. Rose and A. D. Long (2009). "Experimental evolution reveals natural selection on standing genetic variation." Nat Genet **41**(2): 251-257.

ter Linde, J. J., H. Liang, R. W. Davis, H. Y. Steensma, J. P. van Dijken and J. T. Pronk (1999). "Genome-wide transcriptional analysis of aerobic and anaerobic chemostat cultures of Saccharomyces cerevisiae." J Bacteriol **181**(24): 7409-7413.

Thomas, B. J. and R. Rothstein (1989). "Elevated recombination rates in transcriptionally active DNA." Cell **56**(4): 619-630.

Torres, E. M., T. Sokolsky, C. M. Tucker, L. Y. Chan, M. Boselli, M. J. Dunham and A. Amon (2007). "Effects of aneuploidy on cellular physiology and cell division in haploid yeast." Science **317**(5840): 916-924.

Turner, T. L. and P. M. Miller (2012). "Investigating natural variation in Drosophila courtship song by the evolve and resequence approach." Genetics **191**(2): 633-642.

Wilke, C. M. and J. Adams (1992). "Fitness effects of Ty transposition in Saccharomyces cerevisiae." Genetics **131**(1): 31-42.

Wu, J., N. Zhang, A. Hayes, K. Panoutsopoulou and S. G. Oliver (2004). "Global analysis of nutrient control of gene expression in Saccharomyces cerevisiae during growth and starvation." Proc Natl Acad Sci U S A **101**(9): 3148-3153.

Yan, N. (2013). "Structural advances for the major facilitator superfamily (MFS) transporters." Trends Biochem Sci **38**(3): 151-159.

Zhang, E. and T. Ferenci (1999). "OmpF changes and the complexity of Escherichia coli adaptation to prolonged lactose limitation." FEMS Microbiol Lett **176**(2): 395-401.

Ziv, N., N. J. Brandt and D. Gresham (2013). "The use of chemostats in microbial systems biology." J Vis Exp(80).

VITA

Aaron Miller was born in Hannover, New Hampshire. He graduated with a Bachelor of Sciences degree in Biochemistry and Molecular Biology from the University of California at Santa Cruz in 2005. He then moved to the Bay Area to work at the Buck Institute for three years. He then moved north to Seattle to pursue graduate studies in the laboratory of Dr. Maitreya Dunham at the University of Washington, earning his Doctor of Philosophy degree in Genome Sciences in 2015.