

THE EVOLUTION OF GENETIC SYSTEMS: THE INFLUENCE OF RECOMBINATION,
MUTATION RATE, AND MUTATIONAL LOAD

Benjamin Galeota-Sprung

A DISSERTATION

in

Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

Paul Sniegowski, Professor of Biology

Graduate Group Chairperson

Brian D. Gregory, Associate Professor of Biology and Graduate Group Chair

Dissertation Committee

Dustin Brisson, Associate Professor of Biology

Mark Goulian, Charles and William L. Day Distinguished Professor in the Natural Sciences

Joshua Plotkin, Walter H. and Leonore C. Annenberg Professor of the Natural Sciences

Paul Schmidt, Professor of Biology

ACKNOWLEDGMENT

I would first like to thank my advisor, Paul Sniegowski. Paul taught me how to think as a scientist, as an evolutionary biologist, and specifically as an evolutionary biologist interested in mutation rates. He also took a chance on me as someone coming to evolutionary biology with no real background in the area, and always treated my lack of experience and subject matter knowledge most patiently. I can only hope to emulate the level of thoughtfulness that he exhibits—towards the science, and towards others—as I proceed in my career.

The members of my committee, Dustin, Josh, Mark, and Paul all provided important feedback, advice, and different perspectives during the course of my thesis research.

When I first joined the lab, I knew very little about the relevant techniques. Tanya Singh and Mitra Eghbal, at the time both graduate students, were very generous with their time and expertise as I navigated the long path towards competency with basic microbiology skills. I remember well Tanya walking me through a transduction step by step—my first ever genetic manipulation.

I was fortunate to work closely with two committed, responsible, and talented undergraduates, Brooks Martino and Ankur Makani. Brooks carried out much of the work that underpins Chapter 1, tolerating long hours of picking, restreaking, and counting colonies. Ankur worked diligently on a follow-up experiment that we hope will yet bear fruit.

Speaking of committed, responsible, and talented, Breanna Guindon came on board as our research technician just as the yeast work was beginning. As no one had worked with yeast in the lab for quite some time, there was a steep learning curve to master the relevant tools and techniques. Breanna was instrumental to climbing that curve and developing our experimental system, and a large part of the data for Chapters 2 and 3 was gathered by Breanna.

Since Breanna's departure, her role has been admirably assumed by Arlene Garcia, who as I write this Acknowledgment is driving the *C. glabrata* work mentioned in the Conclusion.

I worked closely with Eugene Raynes, a former student of Paul's, on the material in Chapter 1. In fact, I owe the idea to Eugene. Since that work Eugene has been a helpful sounding board for ideas, and also a wonderful conference buddy. I am very thankful to Phil Gerrish. He explained—and re-explained—many population genetics concepts to me, both in the abstract and in relation to my work. I would also like to thank Kayla Peck, who organized a virtual mutation rate journal club, now in its third year, that has helped me connect to more peers and become aware of more work.

Finally, I would like to thank my family, especially my parents, my brother, and my partner Liz, for their love and support.

ABSTRACT

THE EVOLUTION OF GENETIC SYSTEMS: THE INFLUENCE OF RECOMBINATION, MUTATION RATE, AND MUTATIONAL LOAD

Benjamin Galeota-Sprung

Paul Sniegowski

This dissertation investigates the influence of recombination upon the evolution of mutation rates, and the properties of mutational load in evolving populations, using the tools of experimental evolution. It is shown, in Chapter 1, that conjugation inhibits the spread of mutator alleles in an experimental bacterial population. This result confirms previous experimental and theoretical findings that mutator alleles spread by remaining in linkage with beneficial mutations. The exchange of genetic material across individuals disrupts this process, and thereby makes it more difficult for mutator alleles to rise in frequency. This phenomenon likely plays an important role in limiting mutation rates in natural populations.

In Chapter 2, the fitness cost of mismatch repair deficiency is carefully measured in experimental yeast populations, and found to be substantial (>2%). It is shown that this fitness cost is indirect, attributable to a heavy tail of less fit individuals in the distribution of fitness of the mutator population. Separately, the lethal mutation rate of the mismatch-repair-deficient strain is estimated by observing and tracking thousands of budding events of single cells. The reduction in fitness caused by the presence of less-fit individuals in the mutator population and the excess lethal rate of the mutator strain neatly sum to account for the separately measured fitness cost relative to the wild-type strain.

The methods of Chapter 2 are extended in Chapter 3 to produce time-series data from an evolving population. A population of mismatch-repair-deficient yeast is founded from a single cell, in order to begin from a state of no mutational load, and the development of load as the population moves towards mutation-selection balance is measured by estimating the distribution of fitness at several time points. Loads are computed at the early time points, and the methods of approximate Bayesian computation are applied to estimate the deleterious mutation rate and distribution of fitness effects. It is found that the deleterious lethal mutation rate is at least 0.03, and perhaps as high as 0.08, in this strain. These results confirm and augment the findings of Chapter 2, and provide the first-ever experimental demonstration of a population approaching mutation-selection balance.

Table of Contents

ACKNOWLEDGMENT	ii
ABSTRACT	iii
LIST OF TABLES	vii
LIST OF FIGURES	viii
INTRODUCTION	1
1 CHAPTER 1: Conjugation inhibits the spread of mutator alleles in evolving <i>E. coli</i> populations.....	4
1.1 Abstract.....	4
1.2 Introduction.....	4
1.3 Methods.....	5
1.4 Results and Discussion	7
1.5 Conclusions.....	12
2 CHAPTER 2: Measuring and partitioning the mutational load in mismatch-repair-deficient <i>S. cerevisiae</i>	14
2.1 Abstract.....	14
2.2 Introduction.....	14
2.3 Materials and Methods.....	16
2.4 Results.....	20
2.5 Discussion	24
2.6 Supplemental figures and tables	30
3 CHAPTER 3: Load and the distribution of fitness over time in mismatch-repair-deficient <i>S. cerevisiae</i>	37

3.1	Introduction.....	37
3.2	Methods.....	39
3.3	Results.....	41
3.4	Discussion	47
3.5	Supplemental Figures.....	49
4	CONCLUSIONS AND NEXT STEPS.....	50
4.1	Conclusions.....	50
4.2	Predicting mutational load in a non-model organism.....	50
4.3	Summary.....	54
5	REFERENCES.....	55

LIST OF TABLES

Table 2-1 Counts and frequencies of events of interest in the lethal assay	23
Table 2-2 Lethal events by mother / daughter	34
Table 2-3 Lethal events by final microcolony size	35
Table 2-4 Lethal events by size, strain, and mother/daughter	35
Table 2-5 Collected canavanine fluctuation tests for <i>S. cerevisiae</i>	35
Table 2-6 Homopolymeric occurrence in coding sequences for <i>S. cerevisiae</i> and <i>E. coli</i>	36
Table 3-1 Observed mutational load	42
Table 3-2 Estimating s from simulated data.....	44
Table 3-3 Estimating the average selection coefficient.....	45
Table 3-4 Simulations of Day 7 data	47

LIST OF FIGURES

Figure 1-1 <i>mutL</i> - evolution experiments	8
Figure 1-2 Simulations of mutator dynamics in conjugating populations	10
Figure 1-3 <i>mutS</i> - evolution experiments	11
Figure 2-1 Schematic of lethal assay	20
Figure 2-2 Fitness deficit of <i>mmr</i> strain.....	21
Figure 2-3 Distributions of fitness in haploid wild types (A) and <i>mmr</i> mutators (B).....	22
Figure 2-4 Change in <i>mmr</i> fitness disadvantage with ploidy	24
Figure 2-5 Fluctuation tests.....	30
Figure 2-6 Distributions of fitness in diploid strains	31
Figure 2-7 Photographs of two lethal events	32
Figure 2-8 Homopolymeric repeats in URA3	33
Figure 2-9 The lethal and nonlethal loads sum to the competitive fitness difference	34
Figure 3-1 The approach to mutation-selection balance	38
Figure 3-2 The distribution of fitness over time	43
Figure 3-3 Best fits to Day 7 by simulation	46
Figure 3-4 Best fit over time with beneficial mutations	46
Figure 3-5 The association between colony size and fitness	49
Figure 4-1 GC content and HPR occurrence for prokaryotes	52
Figure 4-2 HPR occurrence vs HPR length for selected prokaryotes and yeasts	53
Figure 4-3 <i>S. cerevisiae</i> and <i>C. glabrata</i> are related	54

INTRODUCTION

The central objective of the field of evolutionary biology is to explain the variation in the living world that surrounds us. Variation exists at all levels of organization of living things: between the broad domains of bacteria and eukaryotes; among related species, such as the ~3000 members of the infraorder *Anisoptera*, that is, dragonflies; and in the allelic variation within single populations, the surprising (at the time) amount of which was first revealed by electrophoretic studies of allozymes (Lewontin 1974). A central insight of Darwin (1859)—based only on inference and conjecture but since confirmed by the discovery of DNA as the universal genetic material—is that the distinction between these different sorts of variation is largely a matter of degree, and not of kind.

All such genetic variation in the natural world ultimately stems from mutation. The question, What are the evolutionary forces that determine mutation rates? is thus of fundamental evolutionary importance, and motivates the work described in this Dissertation.

Patterns in mutation rates

Drake (1991) noticed that, in a sampling of bacteria, single-celled eukaryotes, and viruses, the genomic mutation rate—a function of the per-base mutation rate and genome size—is relatively constant (about 0.003). This observation offered the tantalizing prospect that evolutionary forces tend to drive the genomic mutation rate to an optimum, or at least an equilibrium.

A force that tends to drive mutation rates higher is the hitchhiking process, in which an allele that increases the mutation rate—a “mutator” allele—has a relatively higher chance, per generation, of becoming associated with a beneficial mutation (Maynard Smith and Haigh 1974; Chao and Cox 1983; Sniegowski et al 1997). An opposing force, tending to drive mutation rates lower, is that an increase in the mutation rate brings along increased association with mutations of deleterious effect, and indeed, most mutations are expected to be deleterious (Fisher 1930). These two forces share the property that they are indirect, in the sense that the net selection on a modifier of mutation rates results from associations with alleles at other, directly selected, loci.

The observations made by Drake have been sustained in important respects. Among microbes, species distinguished by unusual genetic features, including very large and very small genome size (Sung et al 2012), and the lack of the mismatch repair pathway for replication fidelity (Kucukyildirim et al 2016), have all been shown to possess per-base mutation rates such that their genomic mutation rate is in the neighborhood of 0.003, providing evidence for the wide applicability of “Drake’s rule.”

Yet, once multicellular eukaryotes are added to the picture, the pattern no longer holds. While in microbes the per-base mutation rate scales inversely with genome size, in multicellular eukaryotes the opposite trend is observed: the per base mutation rate scales positively with genome size, a situation requiring new explanations.

The drift barrier hypothesis (Lynch et al 2016) suggests that what explains both patterns is the power of natural selection to reduce mutation rates compared to the power of drift. In general, mutations with selection coefficient $s < 1/N_e$ (where N_e is the effective population size) have dynamics governed more by drift than by selection (Ohta 1973); this is the drift barrier. An allele that reduces the mutation rate will be indirectly selected for, because it tends to be less associated with deleterious mutations than do alternative alleles, only if the magnitude of this indirect benefit exceeds the drift barrier of $1/N_e$.

Because the selective benefit of a reduction in deleterious mutation rate scales with the absolute, and not proportional, reduction in mutation rate, it becomes progressively more challenging to further reduce the mutation rate, relative to the power of drift, as the mutation rate is lowered. This suggests that mutation rates are determined by an interplay between the relative power of drift and selection, and may explain the observed mutation rate patterns in natural species in the following way. Microbes tend to have similar, and very large, effective population sizes; thus they experience a similar power of drift, leading to the similarly low genomic mutation rates first noticed by Drake, and entailing a negative association between mutation rate and genome size. For multicellular eukaryotes, effective population sizes are much lower. As effective population size decreases, the relative power of natural selection to both (1) lower the mutation rate and (2) prevent increases in genome size (*i.e.*, accumulation of junk) are diminished, leading to a positive association between the two.

Mutational load

In a freely recombining population, selection acting against an upwards modifier of the genomic deleterious mutation rate U is approximately $2\Delta U s$ (Leigh 1973), while in an asexual population, it is ΔU (Lynch 2011). The latter figure can be interpreted as the difference in mutational load between two subpopulations with different mutation rates. Mutational load is the reduction in mean fitness of a population due to recurrent deleterious mutation (Muller 1950; Crow 1970; Galeota-Sprung et al 2020). For any given deleterious mutation, mutation-selection balance is reached when the rate of removal by selection equals the rate of creation by mutation. At mutation-selection balance for all loci, the load is equal to the total deleterious mutation rate (Haldane 1937). Because mutation-selection balance, and thus the full realization of load, is reached but slowly (Johnson 1999), when beneficial mutations are relatively common the full power of selection acting on a modifier, ΔU , may be of lesser importance than when beneficial mutations are rare.

Much of the non-neutral sequence diversity in natural populations is likely due to mutational load (*e.g.* Charlesworth 2015). Many proposed solutions to a famously difficult and unsolved problem in evolutionary biology, that of the evolution of sex, depend on the consequences of deleterious mutation (Otto 2009). Mutational load is thus both closely connected to the evolution of mutation rates and also a phenomenon deserving of study in its own right.

Recombination and the evolution of mutation rates

With full linkage, a mutator allele may hitchhike to fixation with just one beneficial

mutation, but with free recombination, it only “receives a hitch” for on average 2 generations (Leigh 1973; Lynch 2011). The exchange and reshuffling of genetic information provided by sex thus makes it more difficult for mutator alleles to spread to high frequency. It may seem counterintuitive then, that asexual organisms such as bacteria have very low mutation rates while obligately sexual organisms tend to have very high mutation rates, but this illustrates the complex interplay of factors and difficulty in asserting causation when considering the fundamental attributes of genetic systems. It is likely that, if the drift barrier hypothesis is correct, the lower limit for selection’s efficacy in reducing mutation rates is set by the effective population size, while recombination prevents mutation rates from rising very much over that lower limit.

Outline of thesis

With these motivations in mind, I undertook two broad strands of work, described in the following three chapters.

Chapter 1 relates an experimental test of the fundamental interplay between recombination and mutation rates. We test the hypothesis that recombination will inhibit the spread of mutator alleles via hitchhiking. To do so, we engineered normally asexual bacteria (*Escherichia coli*) to exchange genetic information via conjugation. We show that, as predicted by theory, recombination makes it more difficult for mutators to hitchhike to fixation.

In **Chapters 2 and 3**, I explore load, deleterious mutation rate, and the fitness effects of deleterious mutations in mutator *Saccharomyces cerevisiae*. In **Chapter 2** I demonstrate that mismatch-repair-deficient (*mmr*) strains have a substantial fitness deficit compared to the wild type, and show that this fitness difference is due to load by (1) measuring the distribution of fitness within each of the two competing strains and (2) independently estimating the lethal mutation rate in an individual cell-based assay. Finally I show that these “nonlethal” and “lethal” components of load sum to the measured fitness difference between the two strains. In **Chapter 3**, I extend the approach developed in Chapter 2—measuring the fitnesses of many individuals in order to estimate the distribution of fitness—to the collection of time-series data from an evolving population. I make further inferences about the deleterious mutation rate in the *mmr* strain, observe the development of mutational load over time, and obtain information about the shape of the distribution of fitness effects of new mutations.

Finally, in the **Conclusion** I describe new directions stemming from the work laid out in this Thesis.

CHAPTER 1: Conjugation inhibits the spread of mutator alleles in evolving *E. coli* populations

1.1 Abstract

Genomic mutation rates can evolve as a result of the indirect selection resulting from genetic association between mutation rate modifier alleles and fitness-affecting alleles at other loci (Sniegowski et al 2000). Numerous experimental studies, corroborated by simulations and analytical theory, have shown that modifier alleles that elevate the genomic mutation rate tend to invade non-recombining populations by hitchhiking with beneficial mutations, yet mutation rates in natural populations of microbes are generally low (Raynes and Sniegowski 2014). Most natural populations are known to undergo recombination, which is predicted to impede the spread of mutator alleles by eroding linkage disequilibrium (Kimura 1967; Leigh 1970; Johnson 1999). Because the role of recombination in mutation rate evolution has received almost no experimental attention (but see Raynes et al 2011), the efficacy of recombination in suppressing mutator allele hitchhiking in real populations remains an open question. Here, we examine selection on mutator alleles in experimental populations of *Escherichia coli* engineered to undergo substantial recombination via conjugation. We show that indirect selection favoring a mutator allele is weakened by recombination, as predicted by theory. Computer simulations indicate that conjugational transfer of the mutator locus inhibits mutator allele hitchhiking more effectively than the transfer of linked beneficial mutations, with the suppressive effect of conjugation declining with genetic distance between the mutator locus and the origin of DNA transfer (*oriT*). In agreement with these simulations, we show experimentally that a second mutator allele, distant from the *oriT*, exhibits frequency dynamics that are essentially unaffected by recombination. We discuss the general relevance of our experiments to the suppression of mutator hitchhiking in natural populations.

1.2 Introduction

Most organisms engage in genetic exchange, either through sex or through forms of horizontal gene transfer (HGT) (Gogarten and Townsend 2005). Surveys of genomic polymorphism suggest that the ratio of recombination rate to mutation rate per base pair is similar for prokaryotes and eukaryotes (Lynch 2007). Here, we investigate whether recombination occurring in bacteria via conjugation can inhibit the spread of mutator allele via hitchhiking.

We evolved populations of Hfr+ (high frequency of recombination) *E. coli* polymorphic for mutator alleles in two different genomic backgrounds: one conferring low and one conferring high levels of conjugation. In both backgrounds, an F plasmid was integrated into the bacterial chromosome along with two additional origins of transfer. In the low-recombination (LR) background, the natural surface exclusion system of the F plasmid minimized conjugation rates. In the high-recombination (HR) background, the surface exclusion protein-coding genes *traS* and *traT* were deleted, elevating conjugation rates by approximately two orders of magnitude.

In conjugation, the probability that any given locus will be transferred via conjugation declines exponentially with its physical distance from an origin of transfer (Smith 1991). In this system, the two extra *oriTs* ensure that much of the genome is accessible to transfer via conjugation. In any given conjugation event, genetic transfer is one-way from donor cell to recipient cell, but as individual cells can act as donors or recipients with equal probability, transfer is effectively two-way at the bulk level.

The mutator alleles that we employed are deletions of *mutL* and *mutS*, key constituents of the mismatch repair pathway in *E. coli*. Both elevate the genomic mutation rate approximately 100-fold, as confirmed by fluctuation test. Fitness tests suggested that *mutS*- has no detectable direct effect on the fitness of its bearer; *mutL*- has no effect on fitness in the HR background but has a deleterious effect on fitness in the LR background. Figures 1-1 and 1-3 (insets) illustrate the bacterial chromosome and positions of the mutator loci used in the experiments, relative to the *oriTs* of the integrated F plasmid.

1.3 Methods

1.3.1 Experimental strains, medium, and propagation conditions.

HR and LR chromosomal backgrounds used in this study were previously described (Winkler and Kao 2012). Deletion cassettes of the mismatch repair genes *mutL* and *mutS*, both carrying insertions of the kanamycin resistance gene (*mutL::kanR* and *mutS::kanR*), were transduced into HR and LR backgrounds using a standard PI transduction protocol. Resistance to kanamycin (at 30 µg/ml) was used as a selectable marker to assay mutator frequency in experimental populations. Nalidixic acid resistance mutations, used as selectable markers for non-mutators, were added to nonmutator HR and LR backgrounds by plating aliquots of each on lysogeny broth (LB) agar plates containing nalidixic acid (30 µg/ml) and isolating resistant mutants. Fluctuation tests conducted using a previously described protocol (Raynes et al 2011) showed that the *mutL*- and *mutS*- mutator alleles both produced ~100-fold increases in the mutation rate to rifampicin resistance, compared to the non-mutator backgrounds. Importantly, previous work indicated no significant difference in genomic mutation rate between HR and LR backgrounds themselves (Winkler and Kao 2012). Fitness competitions, conducted as described below, indicated that *mutS*- has no detectable effect on individual fitness (HR background: mean relative fitness = -0.02, $p = 0.4186$; LR background: mean rel. fit. = -0.03, $p = 0.1891$). Likewise, *mutL*- did not have a significant effect on fitness in the HR background (mean relative fitness (5 competitions) = 0.003, $p = 0.2061$) but was moderately deleterious in the LR background (mean relative fitness (5 competitions) = - 0.09, $p = 0.025$).

All experimental populations were propagated in 50-mL Erlenmeyer flasks, with continuous shaking at 120 rpm and 37C. Low-glucose M9 medium supplemented with 1 ml/L 0.2% of thiamine, 25 mg/L of tryptophan, 0.5 ml/L IM of MgSO₄, and 62.5 ul/L of 20% glucose was used for all experiments except where noted below.

1.3.2 Evolution experiments

To initiate evolution experiments, mutator and nonmutator backgrounds were first inoculated into high-glucose M9 medium (2.5 ml/L 20% glucose) from frozen stock. After

24 hours of growth, cultures were diluted in saline and volumes corresponding to ~1000 cells were used to start five replicate mutator and non-mutator cultures in low-glucose M9. After another 36 hours of growth, appropriate aliquots of saturated cultures of mutator and non-mutator carrying backgrounds were combined to establish 5 replicate populations for each background/mutator locus/starting frequency combination described in the main text. Populations were then propagated by daily 1:100 dilution into 10 ml of fresh low-glucose M9 medium, resulting in ~6.6 generations of growth between transfers.

Mutator frequencies were assayed at intervals by plating population samples onto permissive LB agar plates, allowing for growth into colonies, and then streaking 100 randomly selected colonies onto LB agar + kanamycin plates. The time to fixation was estimated as the first time when mutators reached the frequency of 97% of the population and remained above 97% for the rest of propagation. Populations in which mutators never reached fixation during the experiment were assigned the last day of propagation as the mutator fixation time. Since mutators mostly failed to fix in HR populations, this approach made statistical comparisons between HR and LR populations conservative. Experimental populations were archived at -80 C in 15% glycerol at regular time intervals.

1.3.3 Competitive fitness assays

Relative fitnesses of ancestral and evolved backgrounds were measured in short term overnight competitions. Ancestral backgrounds were inoculated from frozen stock into high-glucose M9 and allowed to grow overnight. Evolved backgrounds were also first inoculated into high-glucose M9 from samples frozen during the course of the experiment and after two hours at 37C, kanamycin (at 30 µg/ml) or nalidixic acid (at 30 µg/ml) were added to cultures to select for only the mutator or nonmutator carrying backgrounds. Note that kanamycin resistance is permanently linked to both mutator alleles. On the other hand, nalidixic acid resistance is usually generated by mutations in the *gyrA* locus (Saenz et al 2003) and could theoretically have been transferred to a mutator-carrying background by conjugation. Assays of resistance frequencies over the course of propagation, however, revealed that individuals carrying both were generally very rare (~1%). After cultures were treated with antibiotics, they were also grown overnight to saturation.

The next day, ~1000 cells from each culture were transferred to fresh low-glucose M9 and allowed to grow for another 24 hours before being diluted 100-fold into 5 (or 10) replicate cultures. Competitions were then allowed to grow overnight in the same conditions as the main propagation experiment. Samples of each culture were plated on permissive LB plates before and after the competition and mutator frequencies were estimated as described above. Relative fitnesses were then calculated from the change in marker frequency during the overnight competition using the standard population genetics formulation for the selection coefficient in a haploid population $s = \frac{1}{t} \ln(p_t(1 - p_0) / p_0(1 - p_t))$ (Crow and Kimura 1970).

1.3.4 Computer simulations

Simulations modeled an asexual population of constant size evolving in discrete, non-overlapping generations. Genomes were modeled as lists of 100 loci with 99 fitness-affecting loci and one mutation rate modifying locus. At the beginning of simulation, a population is composed of 10000 individuals carrying no fitness-affecting mutations, and 5% of the population carries a mutator allele at the mutation rate modifying locus. Simulations continued until the mutators either sweep to fixation or go extinct. The fixation probability P_{fix} was calculated as the fraction of simulation runs in which mutators fixed.

The simulation was carried out in the order reproduction, mutation, and conjugation, each generation. To form the next generation of individuals, every class from the current generation is sampled with replacement with probability proportional to its frequency and relative fitness (a Wright-Fisher model).

After reproduction, each lineage may acquire beneficial and deleterious mutations (but no mutation rate modifying mutations). Beneficial and deleterious mutation counts for each lineage i are randomly drawn from Poisson distributions with means $N_i U_d$ and $N_i U_b$ respectively where N_i is the size of the lineage and U_d and U_b are per-individual mutation rates. Fitness effects of new mutations are then sampled from an exponential distribution with mean $s_d = -0.01$ for deleterious mutations and $s_b = 0.1$ for beneficial mutations. New mutations are assigned to positions in the genome randomly chosen from the 99 positions available for fitness-affecting mutations. Additivity of fitness effects is assumed.

To simulate conjugation, a small number of individuals (determined by the conjugation rate parameter) are first randomly assigned into donor-recipient pairs. The donor then transfers a region of its genome to the recipient, *i.e.*, the region in the recipient genome is overwritten with the genome of the donor. Transfer always begins at the single *oriT* and for simplicity of simulation the size of the region is kept constant.

1.4 Results and Discussion

Despite its inherent fitness cost, the *mutL*- mutator allele was strongly and consistently favored by indirect selection in LR populations (Figure 1). When present at an initial frequency of 50%, *mutL*- spread rapidly through LR populations and approached fixation within 80 generations (Figure 1A). When present at an initial frequency of 5%, *mutL*- was similarly successful in four of five replicate LR populations (Figure 1B); in one replicate, however, *mutL*- frequency declined sharply beginning at around generation 50, consistent with the competing nonmutator (*mutL*+) allele becoming fortuitously associated with one or more strong beneficial mutations (Raynes et al 2012; Gentile et al 2013). Fitness measurements at the end of propagation in four LR populations that fixed for *mutL*- (two each at 50% and 5% starting frequency; see Figure 1C and 1D) showed that the evolved *mutL*- genetic backgrounds in these populations were markedly and significantly fitter than the ancestral *mutL*- background (50% *mutL*- starting frequency: LRa and LRb $p < 0.001$; 5% *mutL*- starting frequency: LRb and LRd $p < 0.001$), consistent with the accumulation of one or more beneficial mutations in association with the *mutL*- allele in these populations. Overall, both *mutL*- frequency dynamics and the accompanying fitness

evolution in LR populations were consistent with the well-understood mechanics of mutator hitchhiking (de Visser 2002; Trobner and Piechocki 1984; Chao and Cox 1983; Thompson et al 2006).

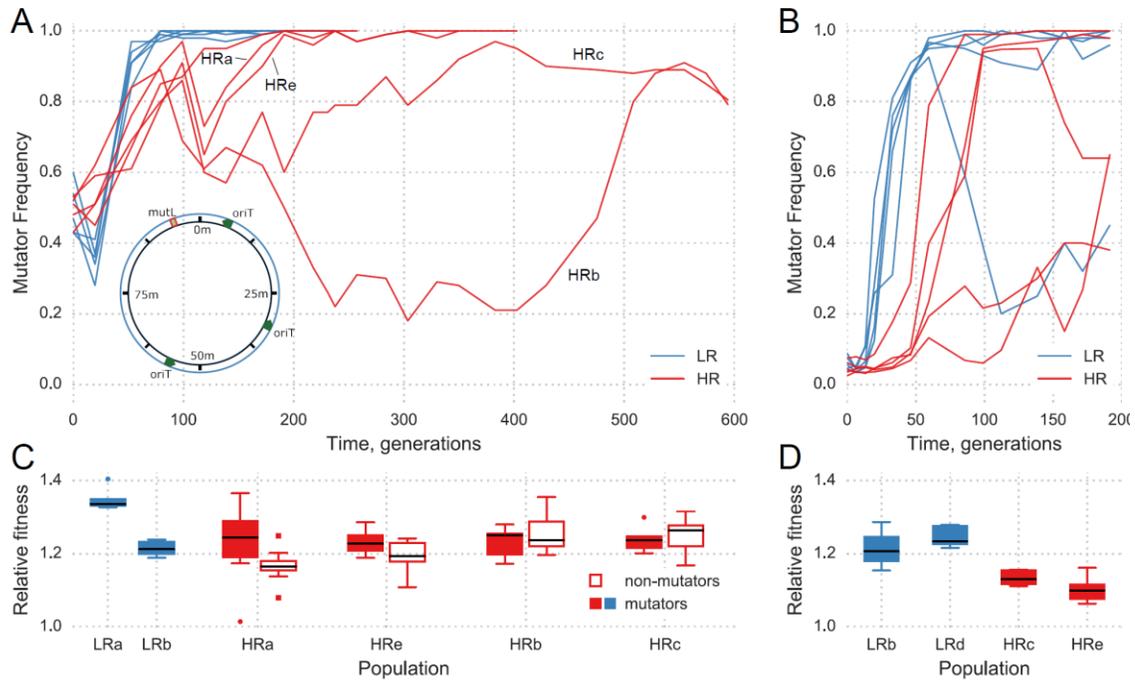


Figure 1-1 *mutL*- evolution experiments

Genetic map of the chromosomal background with the location of the *mutL* locus and three *oriT*. (A) Frequency dynamics of the *mutL*- allele present at an initial frequency of 50% in LR (blue) and HR (red) backgrounds. Replicate populations were initiated by combining isogenic *mutL*- and *mutL*+ populations grown overnight and propagated by daily 1:100 dilutions into fresh medium. Note that the initial decline in *mutL*- frequencies observed in LR populations was likely due to a direct cost of the mutator allele in LR background (B) Frequency dynamics of the *mutL*- allele initially present at the 5% frequency in LR (blue) and HR (red) backgrounds. (C) Relative fitnesses of evolved backgrounds at generation 165 isolated from 50% LR (LRa and LRb) and HR competitions (HRa, HRb, HRc, HRd) in panel A. The relative fitness of each evolved background was estimated from the change in frequency in an overnight competition against the appropriate ancestral background. Each boxplot represents at least 5 (10 for HRa and HRe) replicate fitness. (D) Relative fitnesses of evolved backgrounds at generation 106 isolated from 5% LR (LRb and LRd) and HR competitions (HRc and HRe) in panel B.

In HR populations, in contrast, *mutL*- frequency dynamics were much less predictable across replicates, consistent with recombination weakening the effect of indirect selection. Figure 1-1 shows that *mutL*- frequency trajectories in HR populations were considerably more erratic than in the LR populations and exhibited frequent reversals. Moreover, although periods of *mutL*- increase occurred in all HR populations, *mutL*- fixation was delayed or forestalled in most: when present at an initial frequency of 50%, *mutL*-

required significantly more time ($p = 0.025$) to fix in HR populations than in LR populations. Indeed, only three of five HR populations in which *mutL*⁻ was present at an initial frequency of 50% had reached fixation for *mutL*⁻ by the time propagation was stopped—500 generations after *mutL*⁻ had become fixed in the LR populations. When present at an initial frequency of 5%, moreover, *mutL*⁻ reached fixation in only two HR populations over the course of propagation.

In order to test for evidence that recombination was responsible for the difference in *mutL*⁻ frequency dynamics between the LR and HR populations, we investigated the relationship between fitness gains and frequency in four of the HR populations in which *mutL*⁻ was initially present at a frequency of 50%. To do this, we took advantage of the genetic associations between mutators and kanamycin resistance, and between nonmutators and nalidixic resistance, to isolate subpopulations bearing the *mutL*⁻ and *mutL*⁺ alleles from evolving HR. Short-term competition assays involving these subpopulations revealed that genetic backgrounds bearing the *mutL*⁻ and *mutL*⁺ alleles were both significantly fitter (all *t*-test *p*-values $\ll 0.001$) than their ancestors in all four tested populations at generation 165, when the populations were polymorphic for these alleles (Figure 1C). Moreover, *mutL*⁻ backgrounds were not significantly different in fitness between all four populations examined ($F = 0.177$, $p = 0.72$). However, *mutL*⁺ backgrounds from the two populations in which mutators never approached fixation (HRb and HRc in Figure 1C) were fitter (although the difference was very marginally insignificant: $F = 17.7$, $p = 0.052$) than *mutL*⁺ backgrounds from the two populations (HRa and HRe in Figure 1A, 1C) in which mutators eventually fixed; and there was no statistically significant difference in fitness between the *mutL*⁺ backgrounds isolated from populations that fixed mutators (HRa, HRe) or between *mutL*⁺ backgrounds isolated from populations that never fixed mutators (HRb and HRc): $F = 0.87$, $p = 0.43$). These results suggest that increased conjugational transfer in HR populations allowed the *mutL*⁺ allele to become associated, in some cases, with beneficial mutations that had originated in the *mutL*⁻ genetic background.

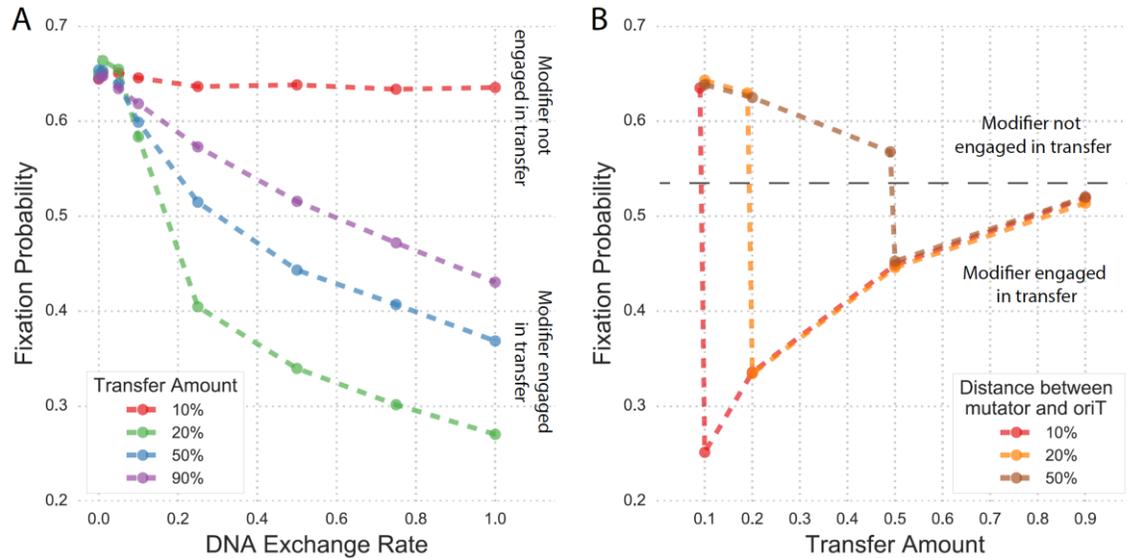


Figure 1-2 Simulations of mutator dynamics in conjugating populations

(A) The effect of conjugation rate on fixation probability (P_{fix}) of a mutator allele. Asexually reproducing Wright-Fisher populations of 10000 individuals were modeled with regular genetic exchange via conjugation. In the model, the genome comprised 100 loci that could acquire both beneficial and deleterious mutations and 1 *oriT*. The mutation rate modifying locus was positioned 20 loci away from the *oriT*. Populations were initiated with the mutator at 5% starting frequency. Parameters: beneficial and deleterious mutation rates = $1e-06$ and $1e-04$ per individual per generation respectively, mean selection coefficients (randomly sampled from an exponential distribution) beneficial, $s_b = 0.1$ and deleterious, $s_d = -0.01$. (B) The effect of the length of transfer region on the inhibitory efficacy of conjugation. Mutator loci were positioned 10%, 20%, and 50% of the genome away from the *oriT*. DNA exchange was set to 50% of the population for all simulations. Other parameters as in panel A. In both panels, fixation probabilities estimated based on 50000 runs of the simulation.

Simulations suggest that the inhibitory effect of recombination depends on whether the mutator locus itself is transferred

There are two ways in which conjugation-mediated recombination can erode linkage disequilibrium between a mutator allele and beneficial mutations: by distributing beneficial mutations from mutator to non-mutator backgrounds, or by replacing the mutator allele with the non-mutator allele on high-fitness backgrounds. Because DNA transfer via conjugation is rare, always starts in the same place in the genome (*oriT*), and transfers smaller regions with higher probability, we hypothesized that the latter process is likely to be more important. In particular, we expected that conjugation-mediated recombination should have the strongest inhibitory effect on the hitchhiking of mutator alleles at loci located downstream from, and close to, the *oriT*: such genetic configurations should facilitate transfer of the non-mutator allele into mutator genetic backgrounds and transfer of the mutator allele into non-mutator genetic backgrounds.

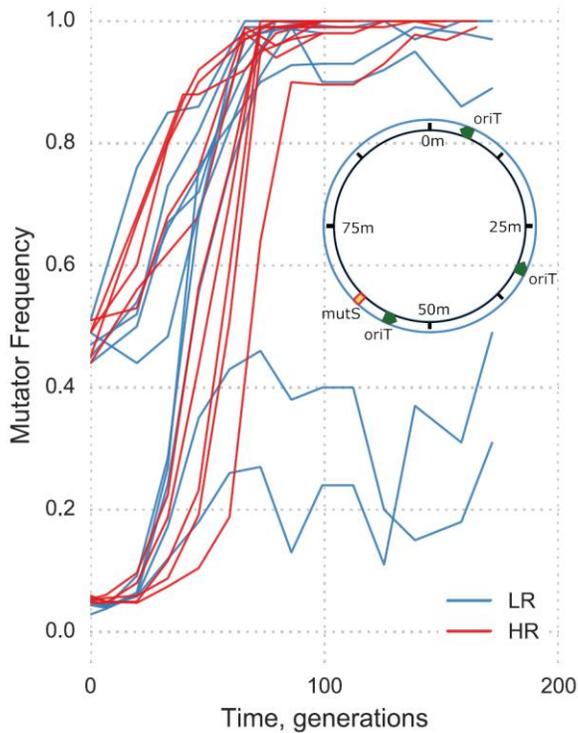


Figure 1-3 *mutS*- evolution experiments

Frequency dynamics of the *mutS*- allele in LR (blue) and HR (red) backgrounds. inset) Genetic map of the chromosomal background with the location of the *mutS* locus.

To investigate this possibility, we simulated mutator frequency dynamics in evolving bacterial populations undergoing regular conjugation and transfer events with different genetic configurations of *oriT*, mutator, and fitness loci subject to deleterious and beneficial mutation. As Figure 1-2A shows, even very frequent conjugation had no apparent effect on mutator hitchhiking when the mutator locus was located just outside the chromosome region transferred. In contrast, the inhibitory effect of conjugation on mutator hitchhiking was maximized when the length of the transfer region was extended to just beyond the mutator locus. Here, non-mutator alleles transferred to mutator backgrounds with only the absolute minimum of their original low-fitness backgrounds while mutator alleles transferred to non-mutator backgrounds with only the absolute minimum of their original high-fitness backgrounds. Notably, when the transfer region was extended even farther, the inhibitory effect of conjugation began to diminish. This suggests that transferring more DNA than necessary to engage the mutator locus can actually benefit the mutator allele by increasing the probability that beneficial mutations remain associated with it after the transfer.

We then investigated the importance of the length of the transfer region in more detail. Here again, conjugation was most effective at inhibiting mutator allele hitchhiking when the region transferred was just sufficiently long to include the mutator locus (Figure 1-2B).

In fact, given that the transfer region included the mutator, the inhibitory effect of conjugation was highest for mutator alleles closest to the *oriT* as these were transferred with the minimum of their genetic background and, in turn, were replaced by the non-mutator alleles with very little of the non-mutator background. Figure 1-2B also illustrates that conjugation-mediated recombination that does not include the mutator locus can nevertheless have an inhibitory effect on mutator dynamics. However, because beneficial mutations can appear anywhere in the genome, substantial inhibition could only be achieved when larger regions of DNA were transferred, which as we have noted before is rather unlikely in nature. Note, for example, that even transfers involving almost half of the genome that did not include the mutator locus (for the mutator locus 50% away from the *oriT*) were not as effective as transfers of much smaller regions that did include the mutator locus.

1.4.1 Recombination via conjugation has no discernible effect on hitchhiking of a *mutS*-mutator allele

Our simulations predicted that the inhibitory effect of conjugation on mutator hitchhiking should weaken with distance of the mutator locus from the *oriT*. To test this prediction, we propagated Hfr+ populations polymorphic for non-mutator and mutator alleles of *mutS*, which is located downstream of, and much farther away from, the *oriT* associated with *mutL* (compare insets of Figures 1-1 and 1-3). In agreement with our expectation, hitchhiking of *mutS*-mutators was essentially unaffected by conjugation (Figure 3). In all HR populations, *mutS* mutators were able to rise quickly in frequency and sweep to fixation. In contrast to the *mutL* experiments, *mutS*-frequency trajectories in HR populations were qualitatively similar to those in LR populations; in particular, we did not observe erratic reversals in frequency dynamics in HR populations. Time to mutator fixation was also not significantly different between HR populations and those LR populations that fixed *mutS*- (50% starting frequency $p = 0.544$, 5% starting frequency $p = 0.145$). Interestingly, however, *mutS*- never reached fixation in two of the LR populations in which its initial frequency was 5%—a result similar to that for one *mutL*-LR population that started at a *mutL*-frequency of 5% (Figure 1B). We speculate that the low initial frequency of the mutator allele, combined with a low conjugation rate, provided scope for clonal interference from beneficial mutations arising on the non-mutator background in these three populations.

1.5 Conclusions

1.5.1 Implications for mutation rate evolution in natural populations

Our experiments and simulations demonstrate that recombination can inhibit and even suppress mutator hitchhiking. A potential limitation of our work is that the rate of recombination exhibited in our experimental *E. coli* populations is likely to be higher than recombination rates in many natural populations. In this respect, it is worth noting that the absolute rate of recombination in a population is not the sole factor affecting hitchhiking: instead, assuming that deleterious mutations are negligible, hitchhiking depends on the strength of selection in favor of beneficial mutations, s_b , relative to the rate of recombination, c , between such mutations and neutral variants: in general,

hitchhiking of a neutral variant linked to a beneficial mutation is favored when $c < s_b$ (Maynard Smith and Haigh 1974). Many experimental populations (including those we have studied here, as confirmed by the fitness gains illustrated in Figures 1-1C and 1-1D) are propagated in novel environments where selection is extremely strong, maximizing the probability of mutator hitchhiking and thus necessitating substantial recombination, as shown here, to suppress hitchhiking. Strong selection and attendant mutator hitchhiking have been inferred in some natural populations (Matic et al 1997; LeClerc et al 1996), but many natural populations are probably so well adapted to their environments that selection is far weaker than in experimental systems; in such cases, much lower rates of recombination should be effective in suppressing mutator hitchhiking. In extremely well adapted populations, moreover, the increased deleterious mutational load associated with a higher genomic mutation rate may play an important role in suppressing mutator alleles, even in the absence of recombination (Wielgoss et al 2013). Thus, while the results that we have presented here demonstrate the powerful effect that recombination can have on mutation rate evolution, this effect, in turn, clearly depends on the relative rates and selective effects of beneficial and deleterious mutations.

CHAPTER 2: Measuring and partitioning the mutational load in mismatch-repair-deficient *S. cerevisiae*

2.1 Abstract

Mutational load is the depression in a population's mean fitness that results from the continual influx of deleterious mutations. Here, we directly estimate the mutational load in a population of haploid *Saccharomyces cerevisiae* that are deficient for mismatch repair. We partition the load in haploids into two components. To estimate the load due to nonlethal mutations, we measure the competitive fitness of hundreds of randomly selected clones from both mismatch-repair-deficient and -proficient populations. Computation of the mean clone fitness for the mismatch-repair-deficient strain permits an estimation of the nonlethal load, and the histogram of fitness provides an interesting visualization of a loaded population. In a separate experiment, in order to estimate the load due to lethal mutations (*i.e.* the lethal mutation rate), we manipulate thousands of individual pairs of mother and daughter cells and track their fates. These two approaches yield point estimates for the two contributors to load, and the addition of these estimates is nearly equal to the separately measured short-term competitive fitness deficit for the mismatch-repair-deficient strain. This correspondence suggests that there is no need to invoke direct fitness effects to explain the fitness difference between mismatch-repair-deficient and -proficient strains. Assays in diploids are consistent with deleterious mutations in diploids tending towards recessivity. These results enhance our understanding of mutational load, a central population genetics concept, and we discuss their implications for the evolution of mutation rates.

2.2 Introduction

An evolving population experiences a continual influx of mutations, the vast majority of which, excluding neutral mutations, are likely to be deleterious (Fisher 1930). A deleterious allele in a haploid population will attain an equilibrium frequency that is the quotient of the mutation rate to that allele and the selection coefficient against it (Danforth 1923). The influx of deleterious mutations causes a depression in the population's mean fitness that is termed the mutational load (Muller 1950), and the load at equilibrium is equal to the deleterious mutation rate (Haldane 1937). Because all populations experience mutation, all populations experience load, and a substantial proportion of the genetic variance for fitness in natural populations is due to mutational load (Charlesworth 2015).

Mutational load is closely connected to the evolution of mutation rates. Consider an asexual population in which there is genetic variation for the mutation rate: within such a population, distinct lineages with differing mutation rates experience differing loads and therefore possess differing mean fitnesses. In this way a downward selective pressure on the mutation rate is realized. This pressure is indirect in the sense that modifiers of the mutation rate are subject to selection without affecting any physiological property immediately related to fitness. The existence of ancient and highly conserved systems for

replication fidelity (including proofreading, mismatch repair, and nucleotide excision repair) attests to the persistence of this selective pressure (Raynes and Sniegowski 2014).

In evolving populations, lineages with higher mutation rates (“mutators”) are continually produced by mutation to any of numerous mutation-rate-affecting loci. In the absence of beneficial mutations, the expected frequency of mutators within a population depends on the increase in the deleterious mutation rate caused by the mutator allele, the rate of mutation from wild type to mutator, and the mean selective effect of newly arising deleterious mutations (Johnson 1999; Desai and Fisher 2011). Investigations of natural and clinical isolates of *Escherichia coli* and other bacteria have shown that mutators of one to two orders of magnitude in strength, often defective in mismatch repair, are present at low but notable frequencies in many populations (Jysum 1960; Gross and Siegel 1981; LeClerc et al. 1996; Matic et al. 1997; Oliver et al. 2000; Denamur et al. 2002; Richardson et al. 2002; Trong et al. 2005; Denamur and Matic 2006; Gould et al. 2007; reviewed in Raynes and Sniegowski 2014). Evolution experiments conducted with *E. coli* have demonstrated that mutators can displace wild types by virtue of their increased access to beneficial mutations (Cox and Gibson 1974; Chao and Cox 1983; Sniegowski et al. 1997; Giraud et al. 2001; Shaver et al. 2002; de Visser and Rozen 2006). Similar findings have been reported for *Saccharomyces cerevisiae* (Thompson et al. 2006; Raynes et al. 2011, 2018). However, in contrast to findings in prokaryotes, mismatch-repair-deficient (henceforth *mmr*) or other types of strong mutators have not been found in natural *S. cerevisiae* populations (but see Bui et al. 2017; Raghavan et al. 2018), though weaker variation for the mutation rate has been detected (Gou et al. 2019). One explanation for this difference could be that *mmr* mutators experience higher load, compared with the wild type, in *S. cerevisiae* than they do in *E. coli*. Indeed, it has been observed by several investigators that haploid *mmr S. cerevisiae* strains decline in frequency in the short term when co-cultured with wild-type strains (Thompson et al. 2006; Raynes et al. 2011, 2018; Bui et al. 2017), even if they eventually out-adapt the wild type. While this short-term deficit of the fitness of *mmr* mutators relative to the wild type has been attributed to increased mutational load in the *mmr* strain, the evidence that this is the case has been mostly circumstantial (but see Wloch et al. 2001) because it is generally difficult to rule out an additional direct fitness effect of any allele thought to cause an indirect fitness effect (Raynes and Sniegowski 2014).

In this work, we establish, by short-term competitive fitness assays and in agreement with prior studies, that an *mmr* haploid *S. cerevisiae* strain is substantially less fit than an otherwise isogenic *MMR+* (*i.e.* wild-type) strain. This fitness difference could be caused solely by load, solely by some direct fitness effect of the *mmr* phenotype, or some combination of the two. We develop separate assays to measure the components of load due to nonlethal and lethal deleterious mutations. To estimate the load caused by nonlethal deleterious mutations, we randomly sampled hundreds of clones from *mmr* and wild-type populations and measured the competitive fitness of each. The resulting histogram of the distribution of fitness of the *mmr* population provides an illustration of the effect of a high mutation rate on population mean fitness. We find that the means of these distributions differ, indicating substantial load for the *mmr* strain, but not fully accounting for the total observed fitness difference between *mmr* and wild-type strains. To

estimate the lethal mutation rate under the two different mutational regimes, we manipulate single cells to track the fate of mother/daughter duos. We show that these two separately measured components of load—due to nonlethal and lethal mutations—approximately sum to the measured fitness difference between the strains; hence we find no reason to suppose a direct fitness effect for *mmr*. Investigations with diploid versions of our strains provide support for this conclusion. We discuss some implications of these findings for continued experimental and theoretical investigations into the evolution of mutation rates.

2.3 Materials and Methods

2.3.1 Data analysis and figure production

Data processing and analysis were performed in R (R Core Team 2019) and RStudio (RStudio Team 2015). Graphical output was produced using the package `ggplot2` (Wickham 2016).

2.3.2 Strains

yJHKII2, a haploid, prototrophic, heterothallic, MAT α , BUD4-corrected, and ymCherry-labeled W303 strain was used as the haploid wild type in all work described in this paper. yJHKIII, labeled with ymCitrine (a variant of YFP) and otherwise isogenic to yJHKII2, was used as the “reference strain” in all haploid fitness competitions. These strains have been previously described (Koschwanez et al. 2013) and were generously provided by the laboratory of Andrew Murray, Harvard University, Cambridge, MA. An *msh2* Δ derivative of yJHKII2, in which the wild-type MSH2 allele was replaced with a kanMX geneticin (G418) resistance cassette (Wach et al. 1994), was used as the haploid *mmr* mutator strain in all work described in this paper. This strain was generously provided by Yevgeniy Raynes of the laboratory of Dan Weinreich, Brown University, Providence, RI and has been previously described (Raynes et al. 2018). The kanMX cassette has been shown to not have a negative effect on growth (Baganz et al. 1997; Goldstein and McCusker 1999).

We constructed diploid versions of each of the three above strains by transforming (Gietz and Schiestl 2007) each with plasmid pRY003, temporarily providing a functional HO locus allowing mating type switching and subsequent mating. pRY003 was a gift from John McCusker (Addgene plasmid #81043; <http://n2t.net/addgene:81043>; RRID:Addgene_81043). The diploid state of resulting isolates was confirmed by (1) ability to produce tetrads after plating to sporulation media; (2) by flow cytometry for total genomic content (following Gerstein et al. 2006); and (3) by the presence of a PCR product for both the MAT α and MAT α loci. The *mmr* diploids would not sporulate, but were confirmed to be diploids by the other two methods.

2.3.3 Growth conditions

The liquid medium for all fitness competitions was synthetic dextrose (SD) minimal medium containing yeast nitrogen base at a concentration of 6.7 g/L and glucose at a concentration of 1.5 g/L (0.15%), supplemented with tetracycline (15 mg/L) and ampicillin (100 mg/L). Fitness competitions were conducted in volumes of 200 μ L in deep

polypropylene 96-well plates (Nunc 260251) sealed with flexible caps (Nunc 276002) and shaken at 1000 rpm with an orbit of 3 mm (Corning LSE 6780-4) at a temperature of 30 °C.

Initial growth in liquid for the lethal event assays was performed in SD as described above but without antibiotics, in flasks shaken at 200 rpm at 30 °C. Growth on agar SD (2% glucose, no antibiotics) plates for the lethal assays took place at room temperature, ~24 °C.

2.3.4 Competitive fitness assays and isolation of clones

Yeast, when grown by batch transfer with glucose as the carbon source, follow a relatively complex cycle of lag, fermentation, and respiration, and fitness benefits “accrued” in one phase (e.g. respiration) may not be “realized” until the next (e.g. the lag following the next transfer) (Li et al. 2018). Therefore we conducted short-term competitive fitness assays between wild-type and *mmr* genotypes in which strains were mixed for one growth cycle prior to measuring frequencies (essentially, following Gallet et al. (2012)). The fitness assays were conducted as follows, with the interval between each consecutive day spanning 24 h. Day 1: wild-type, *mmr*, and the YFP+ reference strain were inoculated from frozen stock into single wells. Day 2: each strain was transferred to a new well with fresh medium, diluting 1/100. Day 3: competing strains were mixed 1:1 by volume and transferred to new wells with fresh medium, diluting 1/100, to create 6 or 8 replicate competitions. Day 4: competitions were transferred to new wells with fresh medium, diluting 1/100, and the frequencies of the competitor and reference strain were assayed by flow cytometry (Guava EasyCyte). Discrimination between strains was performed on the SSC/GRN scatter plot. Day 5: the frequencies of the competitor and reference strain were again assayed by flow cytometry. The population density at the end of a 24-h cycle was $\sim 2 \times 10^7$ cells/mL; the census population size was thus $\sim 4 \times 10^6$ at transfer and $\sim 4 \times 10^4$ just after transfer.

The change in frequencies between Days 4 and 5 was used to calculate a selection coefficient s . Under a continuous model of growth (Crow and Kimura 1970, p. 193)

$$s = \frac{1}{t} \ln \frac{p_t(1-p_0)}{p_0(1-p_t)},$$

from which a relative fitness $w = 1 + s$ follows. The number of generations, t , was assumed to be $\log_2 100$, or ~ 6.64 . p_0 and p_t are the starting and ending frequencies of the genotype being tested (*i.e.* the frequencies at Days 4 and 5 in the above procedure). The resulting selection coefficients represent differences in Malthusian parameter (that is, the log of Wrightian fitness) scaled per generation of growth.

We conducted fitness competitions using this procedure in eight separate blocks, each with multiple replicates as described above. Each block was begun on a different date. For each block, we computed the fitness of the *mmr* strain relative to the fitness of the wild-type strain by subtracting their mutual relative fitnesses to the reference strain. Each block included competitions in both haploid and diploid genotypes. Our final point estimate of the fitness difference between *mmr* and wild-type strains is the mean

difference across all blocks, and the 95% confidence intervals (as shown in Figure 2-1) were computed from the set of point estimates according to the t-distribution.

In two of the eight fitness competition blocks, we randomly sampled individual clones. To do so, we additionally propagated the haploid wild-type and *mmr* strains on Day 3 in addition to mixing them 1:1 with the reference strain. Then, on Day 4, we plated these cultures, diluting appropriately, to YPD agar plates. After sufficient incubation, the resulting colonies were picked by pipet tip into wells containing 200 μ L YPD, grown for 24 h, and frozen down by mixing 1:1 with 30% glycerol and storing at -80°C until needed for fitness assays. The random selection of colonies was ensured by either (1) picking all colonies on a given plate or (2) picking colonies concentrically from a randomly placed point.

Fitness assays for sets of isolated clones were conducted on a single 96-well plate, which allowed us to assay the fitnesses of 88 clones (some wells being reserved for various purposes) or fewer per run. We followed essentially the same procedure as the 5-day competition described above, except that frequencies were estimated at Day 3 and Day 4 instead of Day 4 and Day 5. This modification was made because some clones had such reduced fitnesses that an extra day of growth after mixing 1:1 with the reference strain caused the starting frequency of the clone to depart too greatly from 50%.

The expected variance in measured selection coefficient due to random sampling effects during flow cytometry was computed by means of a simple simulation in which the true frequency of each genotype at the start and end of the fitness competition was replaced by a random binomial variable. Ten thousand replicates of this simulation were run.

2.3.5 Calculation of nonlethal load

Mutational load is classically defined (Bürger 1998) as

$$L = \frac{w_{max} - \bar{w}}{w_{max}}$$

where \bar{w} is the mean fitness of the population and w_{max} is the fitness of the fittest genotype.

We measured all fitnesses relative to a common fluorescent strain, as described above. We define the unloaded fitness of each genotype as equal to 1 and we expect no beneficial mutations to rise to appreciable frequency in the short course of these experiments. Hence, $w_{max} = 1$ and thus

$$L = 1 - w_{max}.$$

We expect measured selection coefficients to be distributed approximately normally around the true value, because of various sources of error including binomial sampling error, drift, instrument noise, environmental perturbations to individual wells (within-batch effects), and among 96-well plates (across-batch effects). To eliminate across-batch effects, for each run we adjusted all measured fitnesses by a constant c such that the mode

fitness is 1 (equivalently, such that the mode selection coefficient is zero). Once this adjustment has been made,

$$L = 1 - \text{mean}(W)$$

where W is the vector of all sampled clone fitnesses, or equivalently

$$L = \text{mean}(S)$$

where S is the vector of all sampled clone selection coefficients. A 95% confidence interval for the nonlethal load was computed by bootstrapping from the measured fitnesses of all sampled clones: for 10,000 replicates, fitnesses were sampled randomly with replacement and the mean computed; from this empirical distribution, the 0.025 and 0.975 quantiles formed the bounds of the confidence interval.

2.3.6 Lethal event assay

Strains were inoculated from frozen stock into 6 mL SD in a flask, and transferred to fresh media after 24 h, diluting 1/100. After an overnight of growth, a streak from the culture was made onto an SD agar plate and five single cells with nascent buds were physically isolated by means of a Zeiss (West Germany) micromanipulating microscope fitted with a Singer Instruments (Somerset, UK) dissecting needle. These cells were periodically checked over the next few hours and the daughters physically separated once developed. These daughters became the founders of microcolonies that were allowed to grow at room temperature for ~20 h, reaching an average size of 23.3 cells (22.8 for the wild type, 23.8 for *mmr*, difference not significant at $p > 0.6$). These microcolonies were then dissected into a gridlike arrangement of single cells (step 1 in Figure 2-1). These cells were then checked at intervals of 1–2 h and daughters separated as soon as possible (step 2 in Figure 2-1). The colonies resulting from these mother/daughter duos were checked at ~24, ~48, and in some cases ~72 and ~96 h after separation. A lethal event was recorded when the growth of a mother or daughter lineage ceased. In such cases cessation of growth was sometimes immediate and sometimes occurred after a few generations. In the latter cases the growth was generally markedly slowed by the first observation. In a few cases, slow but unceasing growth was noted: these are presumed to be cases in which a strongly deleterious mutation occurred, though we stress that this assay was not designed to detect nonlethal deleterious mutations.

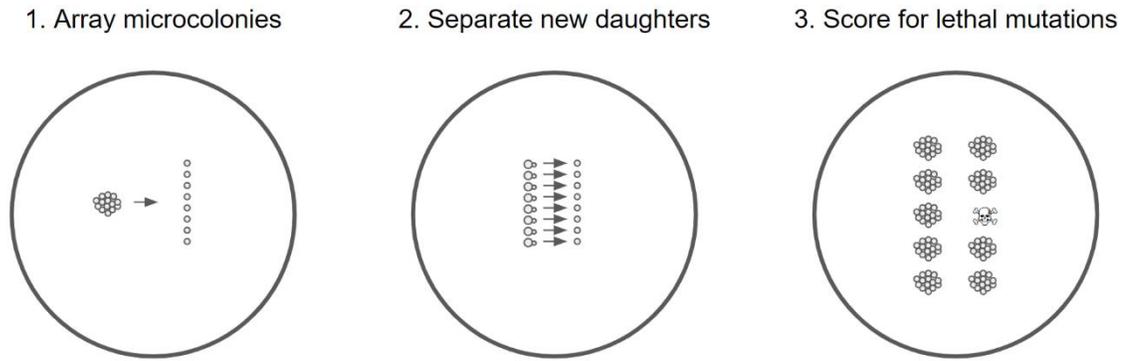


Figure 2-1 Schematic of lethal assay

The arraying in step 1 and separation in step 2 were performed by micromanipulation. Microcolonies in step 1 comprised on average 23 cells, and were founded by new daughter cells that had themselves been isolated by micromanipulation.

As described in the “Discussion” section, the difference in the rate of lethal events between the wild type and *mmr* strains was used as the estimate for the lethal mutation rate for the *mmr* strain. A 95% confidence interval for this difference in rates was computed by the `prop.test` function in R (Newcombe 1998).

2.3.7 Fluctuation assays

To measure the mutation rate to 5-fluoroorotic acid (5-FOA) resistance, we employed the following procedure. Strains of interest were inoculated into 10 mL YPD, grown in flasks with shaking at 30 °C for ~24 h, and then transferred to 30 mL fresh YPD diluting such that ~200 cells were passaged, in replicates of 5. After ~48 h of growth, each replicate was plated without dilution to SD + 5-FOA (1 g/L) agar plates to estimate density and absolute number of resistant, and plated with a 10^{-5} dilution to YPD agar plates to estimate total population density and absolute number. Plates were counted after ~48 h of growth and mutation rates were estimated using the maximum likelihood method of Gerrish (2008). For each round of fluctuation tests, we estimated mutation rates for both wild-type and *mmr* strains simultaneously in order to minimize the influence of any uncontrolled sources of variation.

2.3.8 Homopolymers per gene

The per-base rate of homopolymeric runs of various lengths in *S. cerevisiae* coding regions were computed by a custom Python script. The *S. cerevisiae* S288C reference genome was downloaded from www.yeastgenome.org.

2.4 Results

To confirm that the *mmr* (*msh2Δ*) strain is a mutator, we conducted fluctuation tests using resistance to 5-FOA as the selectable phenotype. Averaged across replicate fluctuation tests, we found a 20.8-fold increase (95% CI: 13.4- to 28.3-fold) in the mutation rate for the *mmr* strain relative to the wild type (Figure 2-5). This is likely an underestimate of the effective genome-wide increase in the mutation rate because *mmr*

mutators have a greatly elevated indel rate for homopolymeric runs (Lang et al. 2013), of which *URA3*, the main locus involved in this fluctuation test, is relatively devoid (Figure 2-8).

2.4.1 Fitness disadvantage of *mmr* compared with wild-type

We competed *mmr* and wild-type strains against a common YFP+ reference strain. Averaged across eight separate blocks of fitness competitions, we found the mutator to be less fit than the wild-type, with an average fitness deficit, expressed as a selection coefficient per generation, of ~2.25% (95% CI: 1.98–2.53%) (Figure 2-2).

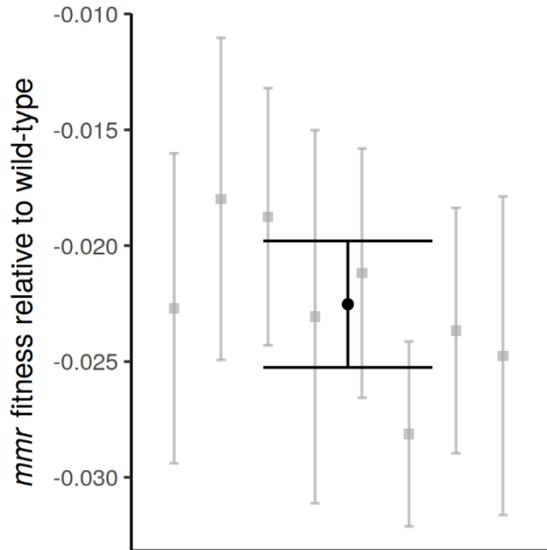


Figure 2-2 Fitness deficit of *mmr* strain

The average competitive fitness deficit (in black, error bars are 95% confidence intervals) of the mutator strain relative to the wild type, expressed as selection coefficient, is ~2.3%. Fitness competitions were conducted in a series of eight blocks, shown in gray. The two strains were not competed directly against each other; within each block, each was competed separately against an otherwise isogenic *MMR+* *YFP+* strain. Each block contained between 6 and 8 replicate competitions.

2.4.2 Estimation of nonlethal load

We randomly sampled individual clones from both *mmr* and wild-type populations and measured the competitive fitness of each clone. The sampled fitness distributions are shown in Figure 2-3. The *mmr* strain's fitness distribution has a prominent left tail of less-fit individuals. We calculated the load as the difference between the mode and the mean fitness; this is ~0 for the wild-type strain and ~1.65% (95% CI: 2.19%–1.08%) for the *mmr* strain. The difference in load between the two strains is significant ($p < 10^{-6}$). The fitness distributions for the *mmr* and wild type are significantly different in shape (Kolmogorov–Smirnov test; $p < 10^{-8}$, while the fitness distribution for the wild-type strain is not

significantly different from a normal distribution with the same mean and variance (Kolmogorov–Smirnov test; $p > 0.05$).

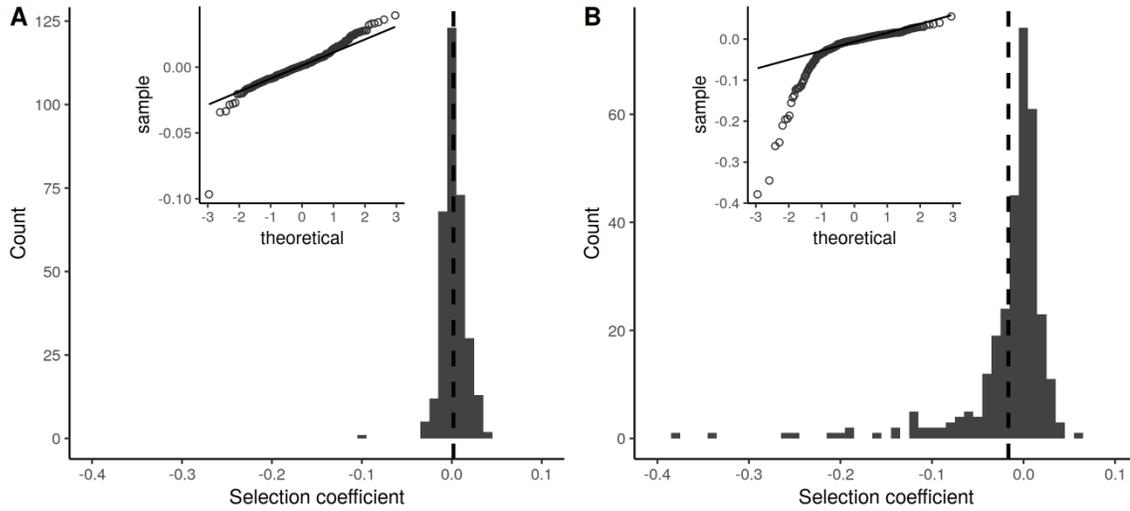


Figure 2-3 Distributions of fitness in haploid wild types (A) and *mmr* mutators (B).

We measured the fitnesses of 327 wild type and 313 *mmr* clones. Fitnesses were measured in competition with an *MMR+ YFP+* reference strain otherwise isogenic to the wild type, as described in Methods. Dashed vertical lines indicate the mean. The load is ~ 0 for wild types; for *mmr* mutators it is $\sim 1.7\%$. QQ plots of the fitness distributions are shown as insets. The two distributions differ significantly (Kolmogorov–Smirnov test, $p < 10^{-8}$).

2.4.3 Estimation of lethal mutation rate

To assay lethal mutation rates, we manipulated single *S. cerevisiae* cells, separating mother/daughter duos and tracking events in which one member of the duo failed to found a colony. The procedure is shown in Figure 2-1 and explicated more fully in the Methods section. Assaying over 2200 duos for each strain, we found a rate of lethal events per newly replicated genome of 0.31% (95% CI: 0.018–0.055%) in the wild-type strain and 0.76% (95% CI: 0.53–1.07%) in the *mmr* strain (Table 2-1). The difference between these rates is 0.44% (95% CI: 0.12%–0.77%). Because the observed rate of lethal events in the wild type is much higher than the expected lethal mutation rate, we take this difference as our estimate of the lethal mutation rate in the *mmr* strain (see the Discussion section for elaboration on this point). Photographs of representative lethal events are shown in Figure 2-7.

Table 2-1 Counts and frequencies of events of interest in the lethal assay

“One lethal” means that the lineage founded by either the mother or daughter cell ceased to grow within the observation period. “One strongly reduced growth” means that either the mother or daughter lineage was observed to grow noticeably slowly. Other events—both members of the duo lethal or strongly reduced growth, or the mother never budding—were not included in the analysis and are not displayed here. The *p*-values reflect the statistical significance of the difference in rates between wild type and *mmr* strains and were obtained by Fisher’s exact test. Note that events are displayed per duo while rates are calculated per individual. “Both OK” means that both mother and daughter cell grew into normal colonies.

Event	Wild-type		<i>mmr</i>		<i>p</i> -value
	Count	Rate	Count	Rate	
Both OK	2235	0.9967	2145	0.9908	0.0006
One lethal	14	0.0031	33	0.0076	0.0049
One strongly reduced growth	1	0.0002	7	0.0016	0.0363

In our assay we followed separated duos that contained a suspected lethal until growth ceased. In some cases growth never ceased, but doubling times were very slow compared with the usual growth rate; such cases were not counted as lethal events but are tallied separately in Table 2-1. We also detected cases in which both members of a duo were lethal, or both showed strongly reduced growth, and also cases in which the mother cell never divided. Because such events probably stemmed from a mutation that occurred prior to the division that created the duo under observation, we excluded these events from our analysis.

2.4.4 Results in diploids

From the haploid strains, we constructed diploid *mmr* and wild-type strains. We calculate the nonlethal load in the *mmr* diploid strain as ~0 for the wild-type diploid strain and 0.30% (95% CI: -0.02 to 0.55%) (Figure 2-6) for the *mmr* strain—substantially less, by about 80%, than the equivalent load in the *mmr* haploid strain (difference significant at $p < 10^{-4}$). The diploid wild-type and *mmr* distributions differ significantly in shape (Kolmogorov–Smirnov test; $p < 10^{-4}$) but do not differ significantly in mean ($p > 0.05$).

We also measured the difference in population fitness between wild-type and *mmr* diploid strains via short-term competitive fitness assays. We found that the *mmr* diploid is less fit than the wild-type diploid by a selection coefficient of ~1.69% (95% CI: 1.40–1.94%) (Figure 2-4). This difference, though larger than expected, is smaller than the fitness difference between wild-type and *mmr* haploid strains by 26% (haploid-to-diploid difference significant at $p < 0.004$).

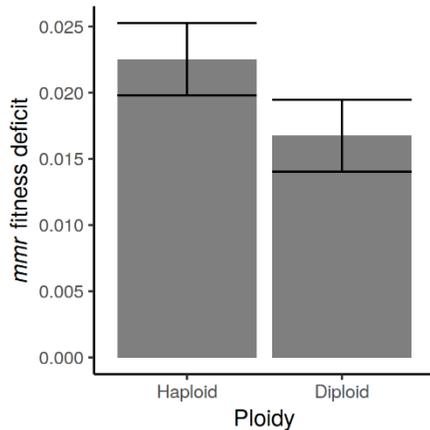


Figure 2-4 Change in *mmr* fitness disadvantage with ploidy

The fitness deficit, relative to the wild type, is 26% lower in diploids than in haploids ($p < 0.004$).

2.5 Discussion

Prior work has shown that, over the short term, haploid *mmr S. cerevisiae* strains decline in frequency when competed with a strain that is wild-type for the mutation rate (Thompson et al. 2006; Raynes et al. 2011, 2018; Bui et al. 2017). Consistent with these findings, we find a fitness disadvantage, expressed as a selection coefficient per generation, of ~2.3% for *mmr* haploids in short-term fitness competitions (Figure 2-2). The magnitude of this selective disadvantage is similar to that in other reports, including Raynes et al. (2011) (2.4% cost), Raynes et al. (2018) (3.3% cost), and Wloch et al. (2001) (4.6% cost, though this is a noncompetitive measure of absolute growth rate).

The deleterious mutations that cause load include both lethal and nonlethal mutations. There is no fundamental theoretical distinction between these two classes of mutation insofar as their contribution to load is concerned: in many population genetic models, all members of an asexual population who are not of the least-loaded class are considered to be doomed (Rice 2002). However, their different manifestations require different experimental techniques. We therefore developed separate approaches to measure these two components of load.

2.5.1 Load due to nonlethal mutations

We measured the short-term competitive fitnesses of 640 randomly selected haploid clones. The histogram and QQ plot for the haploid wild-type populations (Figure 2-3A) suggest that, apart from one less-fit clone, the distribution of fitness for the wild-type strain is essentially normal. The normality of the distribution is consistent with nearly all wild-type clones having the same genotype and thus the same expected fitness, along with many small sources of error in estimation of fitness. One such source of error is drift over the course of the short-term fitness competition. The formula derived by Gallet et al. (2012) suggests that the expected variance in fitness measurement due to drift given our

experimental parameters is $\sim 2 \times 10^{-6}$. A larger source of variance is due to sampling error: in the fitness competitions, we estimate the relative frequencies of the competitors at two time points, sampling ~ 8000 cells per time point. We carried out simulations that suggest that the expected variance due to sampling error is $\sim 2.3 \times 10^{-5}$. These two sources of variance, summed, make up about 20% of the observed variance in selection coefficient. The remainder of the variance probably stems from small-scale environmental variation and other unknown sources of error.

In contrast to the results in wild types, the fitness histogram and QQ plot for the haploid *mmr* strain (Figure 2-3B) are not reflective of a normal distribution. Instead, a prominent left tail of less-fit clones demonstrates the effect of mutational load. The mean selection coefficient is approximately -1.7% , which is the quantification of the reduction in population mean fitness due to nonlethal load. This reduction accounts for a substantial portion ($\sim 75\%$) of the measured competitive fitness difference (Figure 2-2) between the two strains.

Our estimate of the nonlethal load (-1.7%) reflects the average per-generation growth deficit of the mutator subpopulation due to the accumulation of deleterious mutations up to the point of random sampling of clones. We note that this is an estimate of the load at a nonequilibrium state, and is expected to be less than the full load achieved when mutation-selection balance is reached for all loci. Direct observation of mutation-selection equilibrium in a laboratory setting would be challenging because experimental populations rapidly generate adaptive mutations. Our strain-to-strain fitness assays, which found (Figure 2-2) a fitness deficit of $\sim 2.3\%$ for the *mmr* strain relative to the wild type, are likewise reflective of a nonequilibrium state. Since both estimates are derived from the same nonequilibrium populations, they are directly comparable.

Selection coefficients of about the magnitude we observe here cause changes in relative frequency that are extremely rapid in evolutionary terms. For example, a selective deficit of 2% would cause a decline from 50% to 20% frequency in 70 generations. Observing a rapid initial decline of haploid *mmr S. cerevisiae* strains in competition with wild types, some investigators (e.g. Grimberg and Zeyl 2005) have attributed the observed fitness difference to an unknown direct cost (*i.e.* a pleiotropic effect) while others (e.g. Raynes et al. 2018) have assumed that mutational load fully explains the dynamics. The question has remained open, in part because it has been nearly impossible to definitively rule out a direct fitness effect of being *mmr*—any attempt to measure such an effect will be confounded by the indirect fitness effects. By quantifying the indirect fitness effects (*i.e.* load) we seek to determine if a direct effect need be invoked to explain the observed experimental dynamics.

It is not surprising that the nonlethal load accounts for only a portion of the observed fitness difference. The nonlethal load assay relies on the growth of deleterious mutants in order to measure their fitness and thus cannot detect mutants that do not grow, *i.e.* lethal mutations. In order to measure this portion of the load, we designed an assay in which lethal events are directly observed.

2.5.2 Load due to lethal mutations

The lethal mutation rate has long been a matter of interest (e.g. Dobzhansky and Wright 1941). By observing 4435 mother–daughter pairs (duos), we found a rate of lethal events of 0.0076 and 0.0031 for the *mmr* and wild-type strain, respectively.

Our observed wild-type lethal event rate, 0.0031, is on the order of estimates for the genomic mutation rate itself (Drake 1991; Lynch et al. 2008; Zhu et al. 2014; Sharp et al. 2018) and therefore cannot plausibly reflect the rate of lethal mutations. Our interpretation is that, for the wild type, all or most observed lethal events were not caused by genomic mutations and are instead best considered to be nonmutational deaths, perhaps caused by fine-scale environmental fluctuations, experimental manipulation, or other stochastic sources of insult and stress. Observations of relatively high rates of cell death, too high to be due to lethal mutation, are not uncommon. Replicative aging studies of *S. cerevisiae* often observe low but substantial rates of cell death even in very young mother cells (e.g. Chiocchetti et al. 2007; Shcheprova et al. 2008). Rates of cell death on the order of our observed rate for the wild-type strain have also been observed in young bacterial cells (Wang et al. 2010), suggesting that relatively high rates of nonmutational, non-age-related deaths are common among microbes. Our assay design ensured that colonies were young (the oldest cell in a microcolony was on average ~4.3 generations old) and we did not observe a bias in lethal events towards mothers (Table 2-2), so we do not attribute the observed lethal events to senescence. In fact, we observed, across both strains, a bias towards the lethal event occurring in the daughter cell. This difference was not statistically significant ($p = 0.14$), although within the *mmr* strain only we observed 10 lethal events in mothers and 23 in daughters ($p = 0.04$). The observed bias towards daughters dying, if not a sampling effect, could be attributable to smaller daughter cells being relatively more vulnerable to stress. Indeed, increased vulnerability of daughters to environmental sources of stress has been previously reported (Knorre et al. 2010).

An a priori estimation of the wild-type lethal mutation rate can be made as follows. Lang and Murray (2008) conducted careful estimations of the rate of loss-of-function mutations to the *CAN1* locus in a similar background (W303) as the strains used in this work. Multiplying this rate, 1.5×10^{-7} , by the number of genes thought to be essential for viability, ~1100 (Giaever et al. 2002), and accounting for the fact that *CAN1* is longer than the average essential gene gives an expected lethal rate in wild-type haploids of 1.5×10^{-4} . This estimate is on the upper end but within the range of observed rates of accumulation of recessive lethals in several experiments conducted with diploids (Wloch et al. 2001; Hall and Joseph 2010; Nishant et al. 2010; Zhu et al. 2014; Jasmin and Lenormand 2016). Such a rate would suggest that we expected to observe about 0.3 lethal mutations in the wild-type strain in our experiment; we actually observed 14. Therefore, we consider the observed rate of lethal events in the wild type to be an estimate of the rate of nonmutational deaths. The corresponding rate for the *mmr* strain is 0.0076 (difference significant at $p < 0.001$). Making the assumption that nonmutational deaths equally affect both strains, we take the difference between the wild-type and *mmr* lethal event rates, 0.0044 (95% CI: 0.0012–0.0077), as the estimate of the lethal mutation rate in the *mmr* strain. We note that our empirical result is fairly close to the figure obtained by multiplying the wild-type a priori estimate, 1.5×10^{-4} , by the average fold increase in

CANI loss-of-function mutation rate for *mmr* strains in a collection of published reports (44-fold; see Table 2-5). A slightly different methodology, taking the average CANI loss-of-function rate of *mmr* strains from published reports (1.5×10^{-5} ; Table 2-5) and multiplying by 1100 essential genes yields a somewhat higher expected lethal mutation rate of ~0.015.

In many of the lethal events that we observed, growth did not immediately cease but continued for a few generations (Table 2-3) before halting. Limited growth for a few generations after an ultimately lethal mutation occurs has previously been observed (Mortimer 1955). We also observed morphological defects in several lethal events; one such instance is shown in the bottom panel of Figure 2-7. We note that some lethal mutations that we observed could be due to chromosomal losses during mitosis (aneuploidies), but that knocking out *MSH2* has not been observed to greatly increase the rate of such events in haploids (Serero et al. 2014).

2.5.3 Diploid findings

We measured the short-term competitive fitnesses of 573 randomly selected diploid clones. The distribution of fitness for *mmr* diploids (Figure 2-6) suggests that they are substantially less loaded than *mmr* haploids, as would be expected if dominance attenuates the deleterious effects of new mutations. We calculate the nonlethal load in *mmr* diploids as ~0.3%, as opposed to ~1.7% in *mmr* haploids: that is, ~80% of the load has gone away following diploidization. One interpretation of this finding is that deleterious mutations tend to be recessive in diploids. Thus, comparison of the fitness distributions of *mmr* diploids and *mmr* haploids is consistent with a high mutation rate and diploidy shielding the effects of deleterious mutations.

The sampled wild-type diploid clones included more low-fitness individuals than the wild-type haploids (compare Figures 2-6A and 2-3A). We cannot fully explain this observation; one possible explanation is that diploids are more prone than haploids to nondisjunctions causing aneuploid chromosomes, a notion for which there is some experimental support (Sharp et al. 2018).

The relative difference in short-term competitive fitness between wild-type and *mmr* strains is narrowed by 26% in diploids (Figure 2-4). It is somewhat surprising, given that the nonlethal loads are not very different between wild-type and *mmr* diploids, that this figure is not larger. One possibility is that the diploid mutator fixed a deleterious mutation during the process of diploidization, which would account for the discrepancy between the reduction in nonlethal load (82%) and the reduction in total fitness difference (26%) in diploids compared with haploids. Another formal possibility is that diploid mutators have a higher lethal mutation rate than haploid mutators, but we cannot posit a causative mechanism for such an effect.

2.5.4 Considering the two loads together

The total fitness difference between the haploid wild-type and *mmr* strains could be a consequence of greater mutational load for the *mmr* strain, a direct effect of the *msh2Δ* deletion, or a combination of the two. The addition of the lethal and nonlethal loads

($0.0166 + 0.0044 = 0.0210$) is ~7% smaller than the measured fitness difference (0.0225), and the difference is not significant (Figure 2-9). The difference may simply be due to sampling error, or due to systematic underestimation of one or the other of the loads. The nonlethal load may be slightly underestimated because clones were isolated by plating at the beginning of the growth cycle during which competitive fitness was measured. The load may have continued to increase somewhat during this growth cycle.

The broad equivalence of the sum of the loads, on the one hand, and the strain-to-strain competitive fitness, on the other, is consistent with the hypothesis that the total fitness difference is solely due to mutational load. Hence, although we cannot rule out the existence of a small direct fitness effect, these findings suggest that there is no need to invoke direct effects in explaining the fitness difference between the *mmr* and wild-type haploid strains.

The load is equal to the deleterious mutation rate only when the population is in mutation-selection balance. This equilibrium is reached instantly for lethal mutations, quickly for deleterious mutations of large effect, and very slowly for deleterious mutations of slight effect (Johnson 1999). The *mmr* populations in our assays experienced, including the initial process of transformation and growth before frozen storage, about 60 generations of growth, which is enough time to achieve mutation-selection balance for deleterious mutations of relatively large effect, but not for deleterious mutations of slight effect. Hence, our estimate of the total load (2.1–2.3%) should be considered an estimate of the lower limit for the deleterious mutation rate for *mmr* haploids.

2.5.5 Comparison with results in bacteria

Insofar as *S. cerevisiae* and *E. coli* are two model organisms, from different domains of life, with which many evolution experiments have been performed, it is interesting to compare the loads of mismatch repair mutators in both. It appears that in *E. coli* the relative fitness deficit for *mmr* strains is smaller than it is in haploid *S. cerevisiae*. For instance, Shaver et al. (2002) did not detect a fitness difference between *mmr* and wild-type strains, de Visser and Rozen (2006) did not observe an initial decline in *mutS* frequency when that genotype was competed with the wild type at different starting ratios, and Boe et al. (2000) estimated at most a 1% selective disadvantage for *mmr* mutators. In this context it is relevant to note that there are several reports of *mmr* genotypes in natural *E. coli* isolates (LeClerc et al. 1996; Matic et al. 1997; Denamur et al. 2002), as well as in other types of bacteria (Oliver et al. 2000; Richardson et al. 2002; Trong et al. 2005; Gould et al. 2007). In *S. cerevisiae*, in contrast, no functionally *mmr* natural isolates have yet been found (though see Raghavan et al. 2018). Such observations suggest that *E. coli* may be relatively more robust than *S. cerevisiae* to the lack of a functional mismatch repair system. One reason for this difference could be that the genomic mutation rate in *E. coli* is lower than that of *S. cerevisiae* by a factor of about 4 (Lee et al. 2012, Lynch et al. (2008); Zhu et al. 2014; Sharp et al. 2018). If there is a similar absolute difference in deleterious mutation rate, then even if the relative fold increase in the deleterious mutation rate caused by the lack of mismatch repair is also equal in both organisms, the absolute difference in load, which is what controls the evolutionary dynamics, will be larger in *S. cerevisiae* than in *E. coli*. Another possible factor is differences in the spectrum and

genomic substrate of mutations. In both *E. coli* and *S. cerevisiae*, the indel rate is greatly increased in *mmr* lineages, and the rate of indels is strongly elevated in homopolymeric repeats (HPRs). Both the relative increase from wild-type to *mmr* and the absolute indel rate in *mmr* are higher, and scale upwards faster with HPR length, in *S. cerevisiae* than in *E. coli* (Schaaper and Dunn 1991; Tran et al. 1997; Gragg et al. 2002; Lee et al. 2012; Lang et al. 2013). Examining all coding sequences in the *E. coli* and *S. cerevisiae* genomes, we find that there are significantly more HPRs per coding genome, per gene, and per coding base in *S. cerevisiae* than in *E. coli* (Table 2-6). *S. cerevisiae* that are *mmr* are therefore relatively more burdened by indels than are *mmr E. coli* which could account for both the apparent larger fitness difference between *MMR+* and *mmr* and the corresponding apparent contrast in occurrence in natural isolates. We caution that this particular analysis is speculative in nature at this time: one important caveat is that, while this study and others have found large fitness differences between wild-type and *mmr* haploids, *S. cerevisiae* spend most of their time in nature as diploids, in which the fitness deficit of *mmr* lineages might be less severe: classically, equilibrium mutational load is halved in the recessive case compared with the additive, or haploid, case (Kimura et al. 1963). However, while estimates of the rate of outcrossing in *S. cerevisiae* are very low (Ruderfer et al. 2006), the rate of sporulation, which entails a haploid stage, is not known, and evidence of extensive inbreeding and loss of heterozygosity (Peter et al. 2018) suggest that it is relatively frequent. Recessive deleterious mutations may thus be frequently exposed to selection in natural *S. cerevisiae* populations by both the haploid life cycle stage and loss of heterozygosity from inbreeding, suggesting that diploidy may not be as much of a shield for *mmr* lineages as it otherwise would be. A second caveat is that, even if there is no direct fitness effect of *mmr* in haploids, there could be such an effect in diploids, perhaps due to misregulation of the frequency of recombination events (reviewed in Surtees et al. 2004; George and Alani 2012).

2.5.6 Conclusions and future directions

We have found that the indirect fitness effects of strong modifiers for mutation rates are substantial in haploid *S. cerevisiae*, and that it is not necessary to postulate direct fitness effects in order to explain the selective disadvantage of the lack of a functional mismatch repair pathway. This finding is probably most relevant to experimental inquiries of the dynamics of mutation rate evolution in which *S. cerevisiae* is the model organism.

We have also reported findings relevant to fundamental questions about mutational dynamics, including the lethal mutation rate and the relative ratio of lethal and nonlethal deleterious mutations. By sampling the fitnesses of many individuals we have clearly demonstrated mutational load in an *mmr* population, and from the load we are able to estimate a lower limit for the deleterious mutation rate. We sampled hundreds of clones and were able to obtain a clear picture of the left tail of the fitness distribution for the *mmr* strain, but not for the wild-type strain. If the fitnesses of tens of thousands of clones could be measured, much could be learned about load and other evolutionary dynamics at wild-type mutation rates; such experiments may become possible as methods for high-throughput measurements continue to advance.

A limitation of this study is that we captured a snapshot of mutational load at a particular point in time in an evolving population. It would be interesting to observe, at a fine scale, how the distribution of fitnesses changes over time as a population approaches mutation-selection balance, adapts, and experiences other population genetic processes.

2.6 Supplemental figures and tables

2.6.1 Supplemental figures

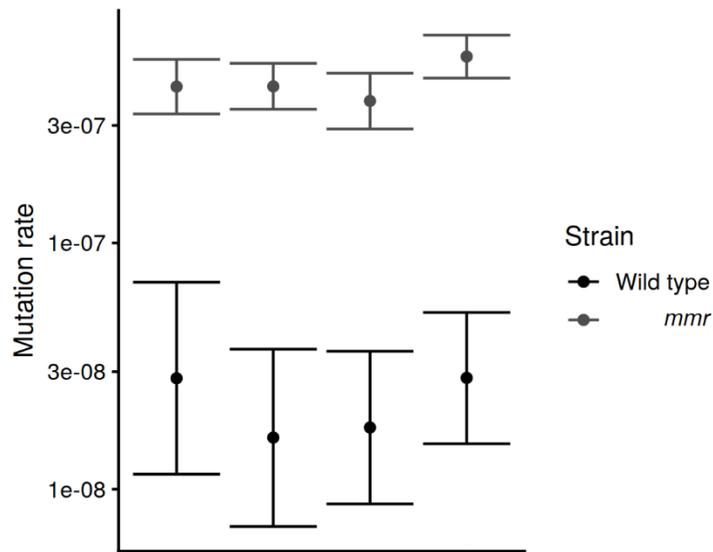


Figure 2-5 Fluctuation tests

Four replicate fluctuation tests, using resistance to 5-fluoroorotic acid, show an average fold increase of 20.8 (95% CI: 13.4 to 28.3) for the *mmr* strain.

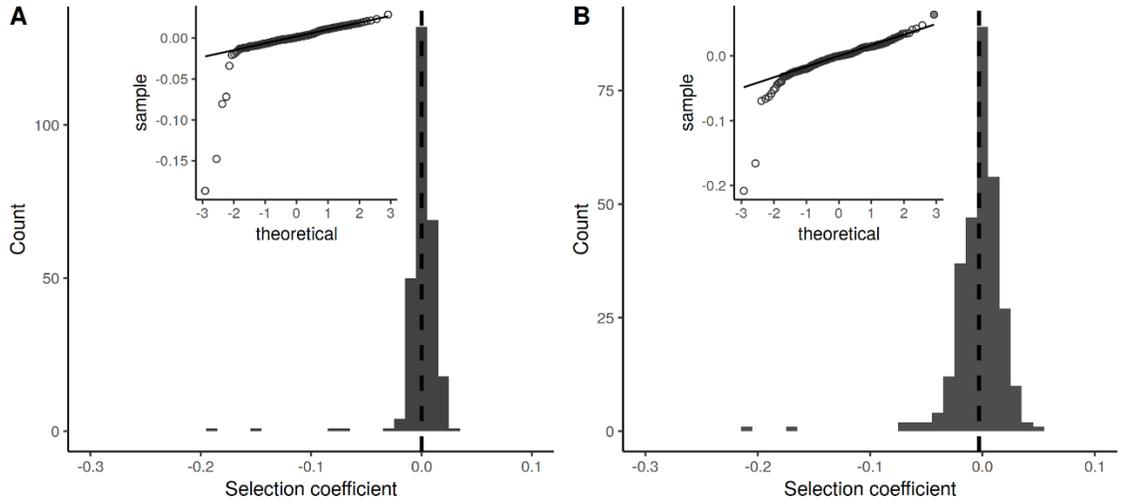


Figure 2-6 Distributions of fitness in diploid strains

Histogram and QQ plot of fitness distributions for diploid strains. Panel **A** shows the results for the wild-type (n=279) strain and panel **B** for the *mmr* strain (n=294). As in Figure 3, the dashed line indicates the mean. The high fitness diploid *mmr* clone shaded in gray was removed from the data set for load calculations and is not shown on the histogram. With removal of this clone the load for the *mmr* diploids is 0.30%; without removal it is 0.28%.

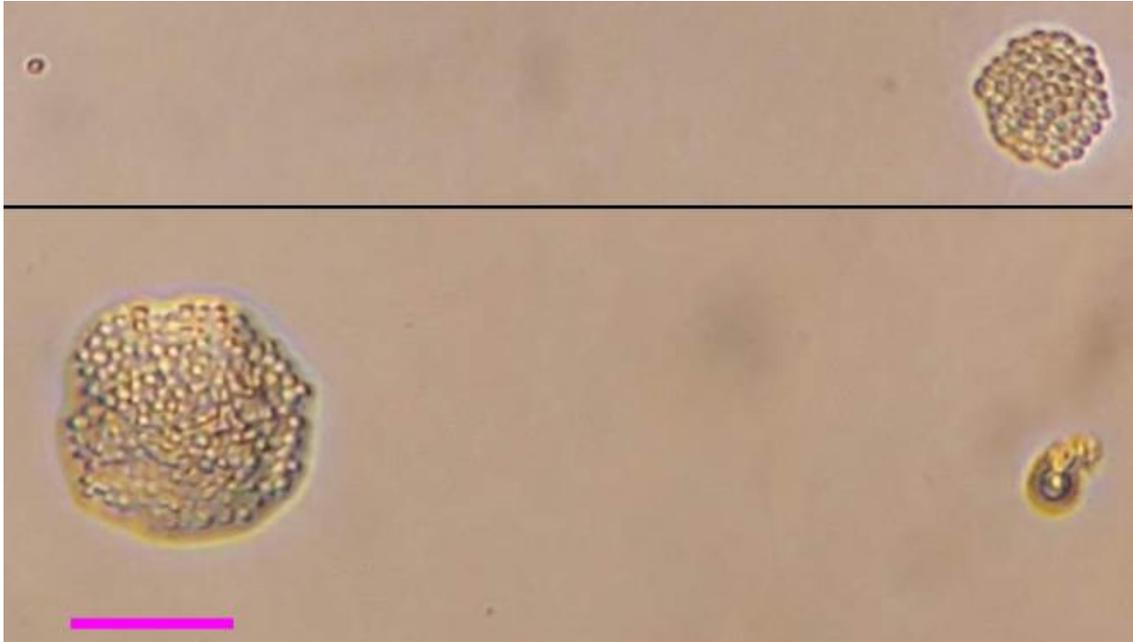


Figure 2-7 Photographs of two lethal events

Examples of two mother / daughter duos that were scored as lethal events. Top: daughter is lethal, photo taken 27 hours after separation. Bottom: mother is lethal, photo taken 29 hours after separation. Scale bar is ~40 microns.

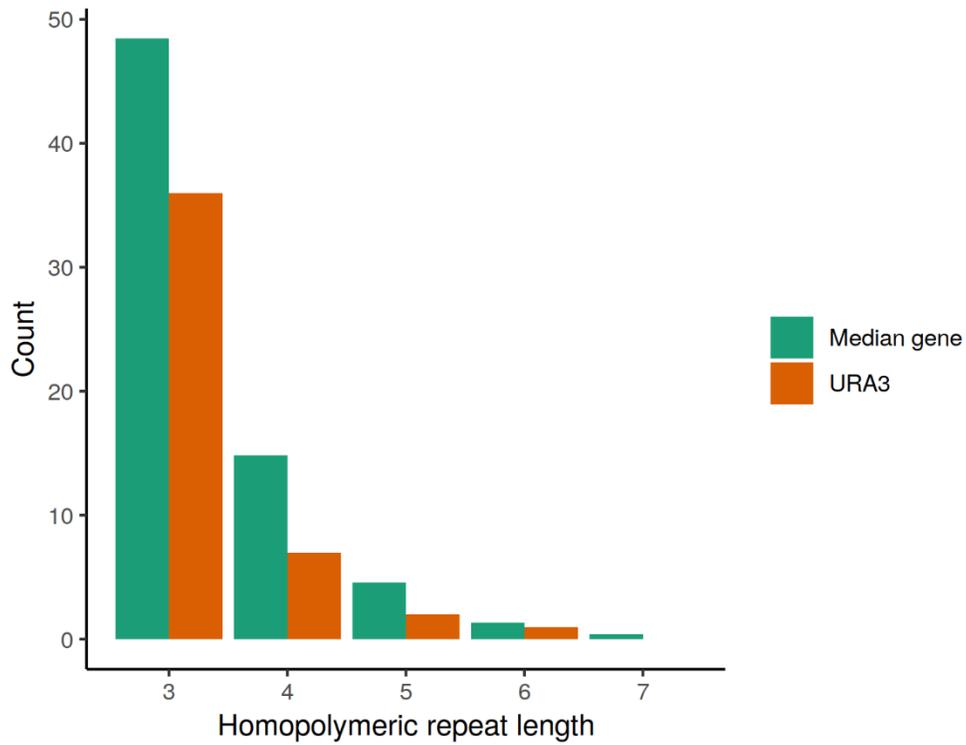


Figure 2-8 Homopolymeric repeats in URA3

URA3 has fewer homopolymeric repeats than the “median gene”, the HPRs per gene as expected by the per-base HPR rate for the entire genome multiplied by the median gene length. This is largely but not entirely driven by URA3 being relatively short.

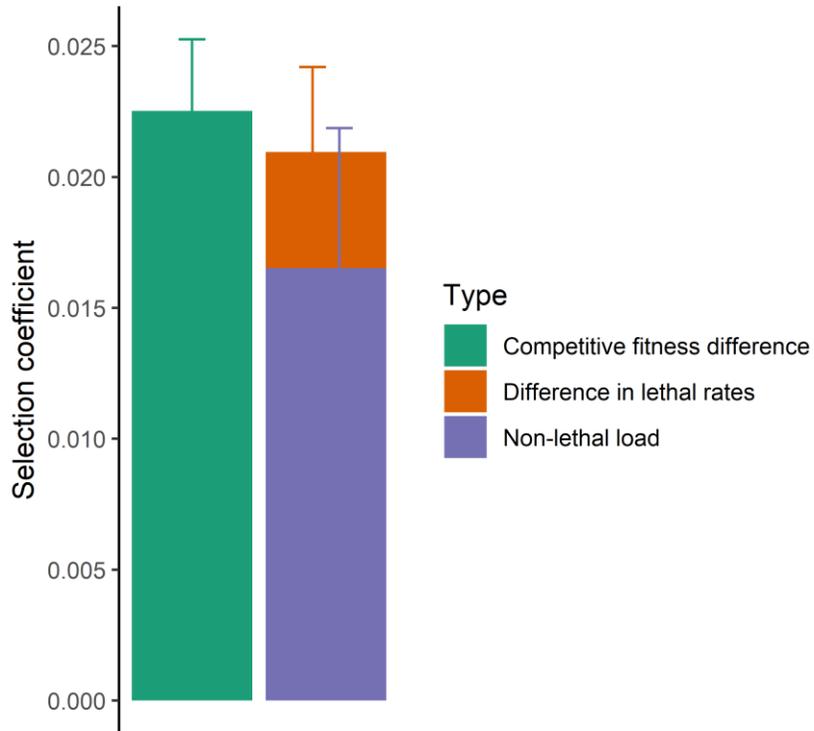


Figure 2-9 The lethal and nonlethal loads sum to the competitive fitness difference

The sum of the nonlethal and lethal loads is ~93% of the measured competitive fitness deficit of the haploid *mmr* strain relative to the wild-type strain. The difference between the two is not significant ($p > 0.2$). The difference between the competitive fitness difference and the nonlethal load alone is significant ($p < 0.02$).

2.6.2 Supplemental tables

Table 2-2 Lethal events by mother / daughter

	Wild type	<i>mmr</i>	Total
Mother	8	10	18
Daughter	6	23	29

Table 2-3 Lethal events by final microcolony size

Final size (number of cells)	Number
1-2	15
3-30	25
31-100	3
100+	4
Total	47

Table 2-4 Lethal events by size, strain, and mother/daughter

Wild-type

Final size	Mother	Daughter	Total
1-2	4	4	8
3-30	4	2	6
31-100	0	0	0
100+	0	0	0
Total	8	6	14

mmr

Final size	Mother	Daughter	Total
1-2	1	6	7
3-30	7	12	19
31-100	1	2	3
100+	1	3	4
Total	10	23	33

Table 2-5 Collected canavanine fluctuation tests for *S. cerevisiae*.

Author and year	Strain	Locus	Genotype	Rate	Fold increase
Lang & Murray 2008	W303	CAN1	WT	1.5E-07	n/a
Zeyl & de Visser 2001	Y55	CAN1	WT	3.2E-07	
Zeyl & de Visser 2001	Y55	CAN1	<i>msh2</i>	1.7E-05	53
Lang et al 2013	W303	CAN1	WT	8.0E-07	
Lang et al 2013	W303	CAN1	<i>msh2</i>	6.7E-06	8
Gammie et al 2007	W303	CAN1	WT	4.8E-07	
Gammie et al 2007	W303	CAN1	<i>msh2</i>	1.5E-05	31
Reenan & Kolodner 1992	SK1	CAN1	WT	4.0E-07	
Reenan & Kolodner 1992	SK1	CAN1	<i>msh2</i>	3.4E-05	85
Marsischky et al 1996	MGD	CAN1	WT	1.0E-07	
Marsischky et al 1996	MGD	CAN1	<i>msh2</i>	4.0E-06	40
Average of above		CAN1	WT	3.8E-07	
Average of above		CAN1	<i>msh2</i>	1.5E-05	44

Table 2-6 Homopolymeric occurrence in coding sequences for *S. cerevisiae* and *E. coli*

p-values are given for the difference in per base rates; “***” indicates a p-value of less than 2.2×10^{-16} .

Homopolymeric repeats in coding sequences

Repeat length	<i>S. cerevisiae</i>			<i>E. coli</i>			<i>S. cerevisiae</i> / <i>E. coli</i> ratio			
	Number	Per gene	Per base	Number	Per gene	Per base	Per coding genome	Per gene	Per base	p-value
3	331178	57.31	4.0E-02	141596	32.830	3.5E-02	2.3	1.7	1.2	***
4	102058	17.66	1.2E-02	35301	8.185	8.7E-03	2.9	2.2	1.4	***
5	31363	5.43	3.8E-03	10895	2.526	2.7E-03	2.9	2.1	1.4	***
6	9287	1.61	1.1E-03	2997	0.695	7.3E-04	3.1	2.3	1.5	***
7	3066	0.53	3.7E-04	564	0.131	1.4E-04	5.4	4.1	2.7	***
8	971	0.17	1.2E-04	97	0.022	2.4E-05	10.0	7.5	5.0	***
9	306	0.05	3.7E-05	9	0.002	2.2E-06	34.0	25.4	17.0	***
≥10	214	0.04	2.6E-05	0	0	0	n/a	n/a	n/a	***

CHAPTER 3: Load and the distribution of fitness over time in mismatch-repair-deficient *S. cerevisiae*

3.1 Introduction

Mutational load is a central concept in population genetics. The contribution of mutational load, relative to other population genetic processes such as balancing selection, in producing the observed allelic variation in natural populations remains a matter of intense interest (Charlesworth 2015), as does the manner in which load is realized in human populations (e.g. Henn et al 2016). Load is also important in conservation biology (e.g. Higgins and Lynch 2001), among other areas.

Evolution experiments conducted in the laboratory have provided insights into diverse topics including the dynamics of adaptation (e.g. Nguyen Ba et al 2019), the evolution of mutation rates (e.g. Sniegowski 1997, Weilgoss 2012), the evolution of sex (e.g. Becks and Agrawal 2011), and the rate, spectrum, and fitness effects of new mutations (e.g. Böndel et al 2019). While fitness declines have often occurred in experiments that seek to remove natural selection, commonly termed mutation-accumulation experiments, the progression of a freely evolving population from an unloaded to a loaded state has not yet been observed and quantified.

In the work described in the previous chapter, mutational load was demonstrated in a population of mismatch-repair-deficient (henceforth *mmr*) haploid *Saccharomyces cerevisiae*, by measuring the fitnesses of many randomly sampled individual clones from a population. This provided a lower bound for an estimate of the deleterious mutation rate in that strain—the estimate being a lower bound because mutation-selection balance had likely not been achieved. In this chapter, I present work that extends this approach—measuring the fitness of many randomly sampled clones—to a population that is evolving over time, with the quantitative goal of leveraging time-series data in order to draw more precise inferences about the nature of the deleterious mutations that are occurring in the population, and the qualitative goal of observing mutational load develop in a relatively large population in which natural selection is operating.

3.1.1 Experimental overview

As in Chapter 2, I have chosen to use a strain with a greatly elevated mutation rate (“mutators”) because at wild-type mutation rates, many times—one or two orders of magnitude—more clones would have to be sampled to reliably measure load, which is not possible technically with the present approach. Because the mutation rate is high, sizable load would be expected to develop simply in the course of constructing a strain, growing it up in order to freeze it, and reviving it (indeed, substantial load was detected under such conditions in the experiment described in Chapter 2). Hence, for the experiment described in this chapter, I founded an evolving population from a single cell—by definition an unloaded state. From this single *mmr S. cerevisiae* cell, a population was obtained, and from the population 528 clones were randomly sampled at four different time points, yielding 2112 sampled clones in total.

3.1.2 Exploratory simulations

A set of three simple simulations, the results of which are depicted in Figure 3-1, illustrates some of the dynamics that we expect to observe. The simulations share the same deleterious mutation rate (henceforth U_{del}), 0.03, but differ according to the average selection coefficient (henceforth \bar{s}) of the distribution of fitness (henceforth DFE) of new deleterious mutations.

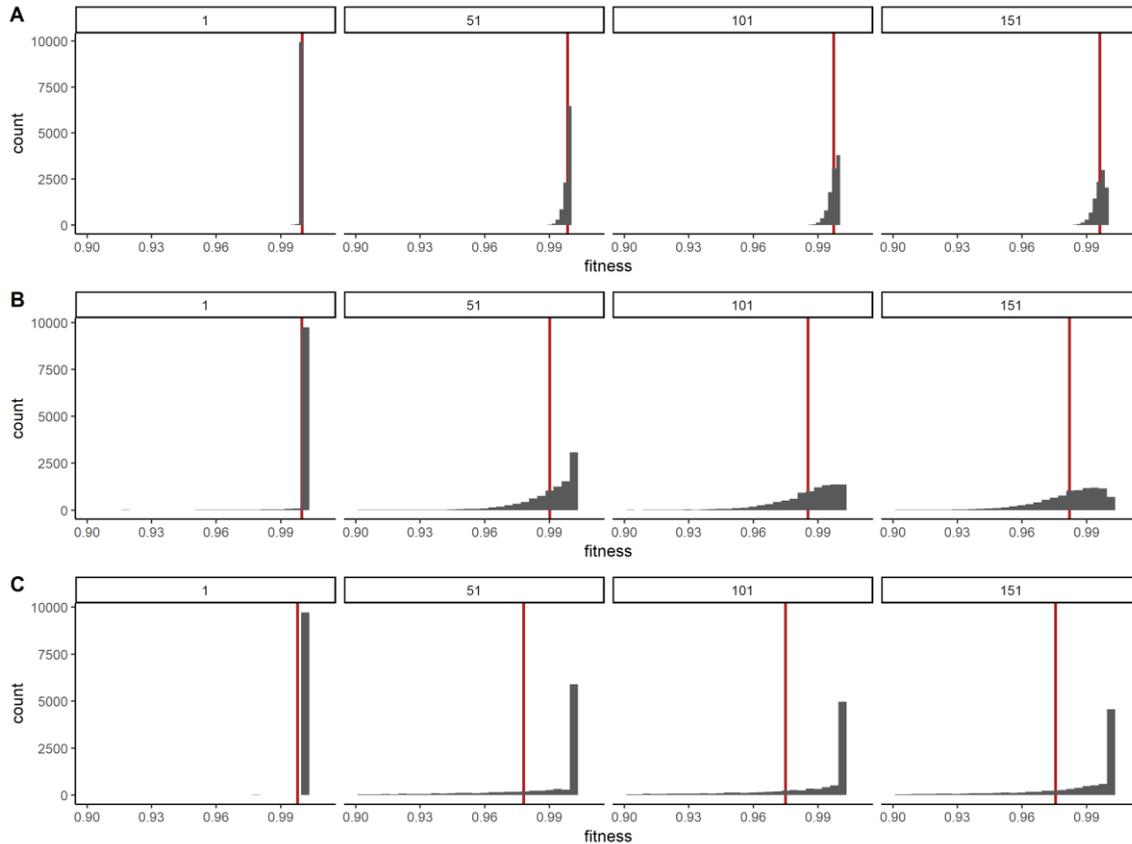


Figure 3-1 The approach to mutation-selection balance

Histograms of fitness of the simulated evolution of three separate populations under deleterious mutation only are shown here. Numbers in boxes are the number of generations elapsed. Initially, all members of the population have fitness 1. The DFE is gamma-distributed and $U_{del} = 0.03$ for all three populations, while \bar{s} varies from 0.001 (**A**) to 0.01 (**B**) to 0.1 (**C**) and with shape parameter always 1. Red lines indicate the mean fitness. Note that the horizontal scale for **A** differs from the horizontal scale for **B** and **C**.

The classical expectation is that all three populations, at mutation selection balance, will have mean fitness $\exp(-U_{del}) = 0.97$ (Gillespie 2010). It is apparent that at any given (non-equilibrium) time point, the three populations differ greatly in both the shape of the fitness distribution and in mean fitness (illustrated by the red line). Consistent with

theoretical expectations (Johnson 1999), the equilibrium is approached much more quickly for higher \bar{s} .

These simulations suggest that the influences of mutations possessing selection coefficients of relatively high magnitude are expressed quite rapidly on the distribution of fitness, on time scales of 50 to 100 generations. Because, as discussed at length in Chapter 2, *mmr* yeast experience a very high indel rate, and indels tend to cause loss-of-function mutations, it is reasonable to assume that such mutations will be observed.

In these simulations, different parameters of the DFE yielded different distributions of fitness. In the experiment described in this chapter, we are largely concerned with the inverse problem—inferring properties of the DFE from measurements of the distribution of fitness over time.

3.2 Methods

3.2.1 Media and growth conditions

The liquid media for population evolution and all fitness competitions was synthetic dextrose (SD) minimal media containing yeast nitrogen base at a concentration of 6.7 g/L and glucose at a concentration of 1.5 g/L (0.15%), supplemented with tetracycline (15 mg/L) and ampicillin (100 mg/L). Evolution of populations, and subsequent fitness competitions, were conducted in volumes of 200 μ l in deep polypropylene 96-well plates (Nunc 260251) sealed with flexible caps (Nunc 276002) and shaken at 1000 rpm with an orbit of 3mm (Corning LSE 6780-4) at a temperature of 30 C. When clones were isolated, they were grown up in YPD medium (1% yeast extract, 2% peptone, and 2% glucose) before freezing.

3.2.2 Strains and pre-evolution

yJHK112, a haploid, prototrophic, heterothallic, MATa, BUD4-corrected, ymCherry-labeled W303 strain that is wild-type for the mutation rate, was transferred by 1/100 serial dilution for 57 days, approximately 380 generations, in an attempt to fix easily-accessible beneficial mutations of large effect. Ploidy was checked at this time and diploids were not detected. A clone isolated from this evolved population was stored as strain YPS3654. From YPS3654, two strains were derived by transformation. A fluorescent strain for use in fitness competitions, YPS3672, was created by transforming a PGK1-GFP-natMX cassette (described in Deschaine et al 2018), generously provided by Helen Murphy. A mismatch repair-deficient (*mmr*) strain was created by transformation with an *msh2::kanMX* deletion cassette. Transformations were confirmed by PCR and elevation of the mutation rate was confirmed by fluctuation test. The *mmr* strain, YPS3660, was used as the founder strain for the experiment.

3.2.3 Evolution experiment

A mutator population was inoculated from a single cell according to the following procedure. From an overnight culture of strain YPS3660, a single cell was isolated on an agar plate by micromanipulation. The cell selected to be isolated was relatively small, to

avoid choosing an older and possibly senescing cell, and visibly possessing a bud, to avoid choosing a dead cell. An agar area surrounding and including the single cell was cut out with a sterile knife and deposited into a 50 mL flask containing 10 mL minimal medium, and placed at 30 C with shaking at 200 rpm. 24h after inoculation (“Day 1”), 1.5 mL of the culture was plated without dilution to several agar plates so that there were approximately 30 colonies per plate, the low number per plate ensuring adequate physical separation during picking. Colonies were allowed to grow for two days and then picked to wells of 200ul YPD on 96-well plates. These clones were grown for 24 h and then frozen 1:1 with 30% glycerol, and thenceforth defrosted when needed for fitness competitions.

48h after inoculation, the culture was diluted 1/100 into 200ul fresh media in a 96-well plate. Every 24h thereafter, the culture was diluted 1/100 into 200 ul fresh media in a 96-well plate.

Clones were isolated on Days 7, 15, 22, by a similar procedure as that described for Day 1, except that dilution was needed to ensure 30-40 colonies per plate. At each time point, 528 clones were isolated. The random sampling of clones was ensured by either (1) picking all colonies on a given agar plate or (2) picking concentrically from a randomly placed mark.

The effective population size for most of the evolution, calculated as the product of the number of cells transferred and the number of generations per transfer cycle, was approximately 3×10^5 .

3.2.4 Fitness assays

Fitness competitions were conducted according to the following four-day procedure. 0h: a plate of frozen clones was defrosted and 2ul from each well was removed to a 96-well plate with 200 ul minimal medium. Also inoculated into wells on the same plate were YPS3672 (GFP+) and YPS3654 (ancestral to the mutator). At 24h, all wells were transferred into fresh minimal medium, diluting 1/100. At 48h, the cultures for each clone were mixed 1:1 with the GFP+ strain and transferred to fresh minimal medium, and the relative frequencies of the two strains measured by flow cytometry. Finally, at 72h, the relative frequencies of the two strains were again assayed by flow cytometry.

Selection coefficients were computed according to the formula

$$s = (1/t) \ln \frac{p_t(1-p_0)}{p_0(1-p_t)} \quad (\text{Crow and Kimura 1970}).$$

This procedure, carried out for one 96-well plate, resulted in competitive fitness measurements for sets of either 88 or 82 clones. Either 6 or 12 replicate “reference competitions” (YPS3672 vs YPS3654) were included on the same plate with each run. For each run, the mean selection coefficient from the set of reference competitions was subtracted from the selection coefficient for each clone to create an adjusted selection coefficient.

Competitions were run in blocks, a block consisting of multiple 96-well plates (*i.e.* multiple sets of 82 or 88 clones) begun on the same day. Each block included sets of

clones from at least two different time points. Batch effects of competition blocks were observed. These were removed, to the extent possible by linear regression, prior to final analysis of fitness data.

Because of technical errors occasionally affecting single wells, portions of a 96-well plate, or an entire plate, the number of fitness measurements per clone varies somewhat. The mean number of fitness measurements per clone is 3.8.

3.2.5 Simulations and search strategy

Individual-based simulations employing a Wright-Fisher model with selection were implemented in the Julia programming language. These simulations were conducted at constant population size, with the effective population size matching that of the experiment. Mutations were drawn randomly from a distribution of fitness, parameterized as described in Results.

An Approximate Bayesian Computation (ABC)-style approach (Beaumont 2010) was employed to use simulations to make inferences about the underlying rate and DFE of new mutations. Parameters were drawn from a prior distribution (uniform on some interval) and a simulation conducted for each set of parameters. The *p*-value of an Anderson-Darling test was used as an inverse distance measure between the simulated and observed distribution of fitness. An appropriate amount of measurement noise for fitness was included in the simulations.

3.3 Results

3.3.1 Colony morphology and “missing” clones

As described in the Methods, in order to randomly sample individual genotypes, at each time point >500 colonies were picked after plating and subsequent growth. At the time of picking, in each case where a given colony appeared smaller than most others on its plate, a note was made. Plotting of fitness data along with colony size (Figure 3-5) reveals a clear association between smaller-than-usual colony size and reduced fitness. This observation suggests that, in general, mutations that are deleterious in liquid culture are often deleterious for growth on agar as well.

Not every clone that was isolated produced a fitness measurement. In total, from the first three time points, 16 clones did not grow well enough after defrosting to allow for competitive fitness measurements: 3 clones from Day 1, 6 clones from Day 7, and 7 clones from Day 15. As the difference in occurrence across days is not significant according to a chi-squared test, these clones were not included in the analysis.

3.3.2 Observed fitness over time and loads

The sampled fitness distributions over time are shown in Figure 3-2. The mean fitness declines from Day 1 to Day 7, and thereafter rises. Mean fitnesses and loads are displayed in Table 3-1. The initial decline in fitness from Day 1 to Day 7 is attributed to the accumulation of deleterious mutations; the rise thereafter is attributed to adaptive mutation. The measured load at Day 1 (~generation 10) was 0.3%, and increased to 2.7% at

Day 7 (~generation 50), using the mode fitness at Day 1 at the max fitness. Across repeated measurements of the same clone, the standard deviation in fitness was about 0.01. This accounts for about half of the variance in fitness measured on Day 1, and <10% of the variance in fitness for the subsequent time points.

At equilibrium mean fitness is $\exp(-U_{del})$. Thus, the Day 7 estimate of the load, 2.7%, is a lower bound for an estimate of the deleterious mutation rate.

Table 3-1 Observed mutational load

Loads are calculated as $(w_{max} - \bar{w})/w_{max}$ with w_{max} equal to the mode fitness on Day 1. By Day 15, adaptation has confounded the load measurement. For this reason load is not calculated for Day 22.

	Day 1	Day 7	Day 15	Day 22
Mean fitness	0.997	0.973	0.981	1.01
Load	0.3%	2.7%	1.9%	n/a

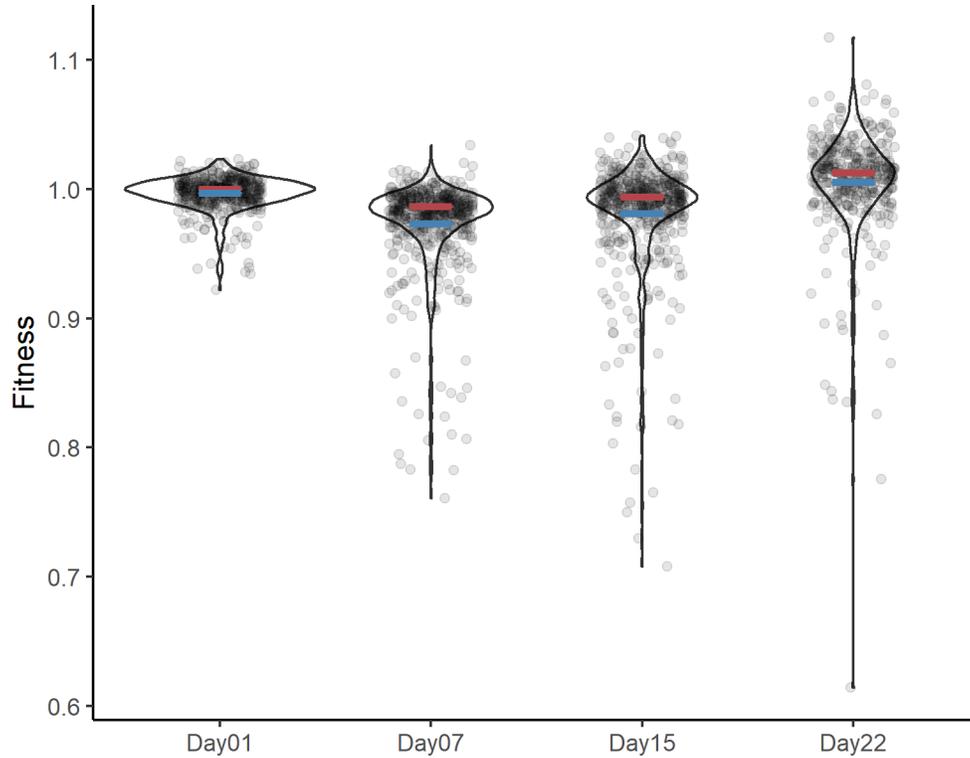


Figure 3-2 The distribution of fitness over time

Violin and dot plots showing the distribution of fitnesses of the sampled clones at each time point. Mode fitness is indicated by the red bar and mean fitness is indicated by the blue bar. Fitnesses have been scaled so that the mode fitness on Day 1 is 1.0.

3.3.3 Estimation of average selection coefficient

Assuming that a population is very close to equilibrium, a good estimator of the deleterious mutation rate is

$$\hat{U}_{del} = -\log(\text{mean}(W)),$$

where W is the vector of all fitnesses. Supposing that all mutations have the same selection coefficient and that there is no epistasis, and letting X be a random variable that specifies the number of deleterious mutations per individual, fitness is $\exp(-Xs)$. At equilibrium, $X \sim \text{Poisson}(U_{del}/s)$ (Haigh 1978), and hence

$$\text{Var}(-\log(W)) = \text{Var}(Xs) = U_{del}s.$$

If there are instead n classes of mutation, each with selection coefficient s_i , and each with proportional mutation rate p_i such that $\sum p_i = 1$ and $\sum p_i U_i = U_{del}$, then at equilibrium

the number of mutations of each class per individual is independently distributed as $X_i \sim \text{Poisson}(p_i U/s_i)$ (Johnson 1999). Then

$$W = \prod \exp X_i s_i$$

$$-\log W = \sum X_i s_i$$

$$\text{Var}(-\log W) = \sum s_i p_i U_{del} = \bar{s} U_{del}$$

where \bar{s} is the expectation of s . This suggests the following estimator for the average selection coefficient:

$$\hat{s} = \text{Var}(-\log W) / -\log(\text{mean}(W))$$

It turns out that this estimator performs surprisingly well even for nonequilibrium populations. Table 3-2 shows these estimators applied to the exploratory simulations summarized in Figure 3-1. Although derived under the assumptions of equilibrium, which is violated in the simulation data—notably, in some cases the populations are quite far from equilibrium—the estimates for \bar{s} are nevertheless always within a factor of 2 of the true value.

Table 3-2 Estimating \bar{s} from simulated data

Examples of the estimators $\hat{U}_{del} = -\log(\text{mean}(W))$ and $\hat{s} = \text{Var}(-\log(W)) / -\log(\text{mean}(W))$ for the three simulated populations depicted in Figure 3-1. \hat{s} is quite robust to nonequilibrium conditions, in these examples always falling within a factor of 2 of the true value even when the populations are far from equilibrium.

True U_{del}	True \bar{s}	Generation	\hat{U}_{del}	\hat{s}
0.03	0.001	1	0.000031	0.0019
0.03	0.001	51	0.0015	0.0019
0.03	0.001	101	0.0028	0.0020
0.03	0.001	151	0.0039	0.0018
0.03	0.001	201	0.0050	0.0018
0.03	0.01	1	0.00025	0.020
0.03	0.01	51	0.0098	0.015
0.03	0.01	101	0.015	0.014
0.03	0.01	151	0.018	0.013
0.03	0.01	201	0.020	0.012
0.03	0.1	1	0.0022	0.14
0.03	0.1	51	0.022	0.093
0.03	0.1	101	0.025	0.088
0.03	0.1	151	0.025	0.080
0.03	0.1	201	0.027	0.078

The application of this estimator to the experimental data is shown below (Table 3-3) for Day 7 only, the later time points being noticeably affected by beneficial mutations.

Table 3-3 Estimating the average selection coefficient

Estimates for U_{del} and \bar{s} for the observed data following the same formulae as employed in Table 3-2. The expected variance due to measurement noise was subtracted from $\text{Var}(W)$ before final computation.

	\hat{U}_{del}	$\hat{\bar{s}}$
Day 7	0.027	0.053

Results from simulation suggest that this method tends to underestimate U_{del} and overestimate \bar{s} for nonequilibrium populations. This parallels the estimates obtained for the same parameters via the classic Bateman-Mukai method for mutation-accumulation data (Bateman 1959; Mukai 1964; reviewed in Keightley and Eyre-Walker 1999), which are subject to the same consideration. Thus, the estimates in Table 3-3 are best interpreted as a lower limit for U_{del} and an upper limit for \bar{s} .

3.3.4 Analysis by simulation and ABC

The load analysis above suggests a lower bound for U_{del} of ~ 0.03 and an upper bound for \bar{s} of ~ 0.05 . In order to infer further properties of the deleterious DFE, we employed a simulation-based ABC-like approach. The DFE was parameterized as a gamma distribution, and simulations over a wide range of parameters (rate, mean effect size, and shape, with the scale parameter constrained by mean and shape) were carried out. As a measure of fit between observed data and simulated data, the p -value of an Anderson-Darling test was used. 1280 combinations of parameters ($16 \times 10 \times 8$) were simulated over a wide range of values. All p -values obtained were $< 1e-04$, suggesting poor fits of simulated to observed data. Inspection of the simulated distributions revealed that either the observed downward shift in mode (compare Day 1 to Day 7 in Figure 3-2) or the development of the heavy left tail could be recapitulated in simulation, but not both, with a single set of parameters.

Therefore a further 3644 simulations in which the deleterious DFE was a mixture of two separate gamma distributions were carried out. The top two fits are shown in Figure 3-3 and all parameter combinations yielding p -values greater than 0.05 are shown in Table 3-4. The simulations suggest that a DFE given by a mixture of two gammas, both with relatively high mutation rates, the combined rate being about $U_{del} = 0.08$, provides a good fit to the observed data. The first distribution has a low mean and high shape parameter, while the second (with a slightly lower mutation rate) has a high mean and low shape parameter. In general, the deleterious DFE parameter values that contributed to simulations with p -values > 0.05 were similar to one another.

Consistent with expectations, the simulation-based estimates for U_{del} and \bar{s} , 0.08 and 0.03 respectively, for the parameters with the highest p-value, are higher and lower respectively than the estimates derived in Table 3-3 above.

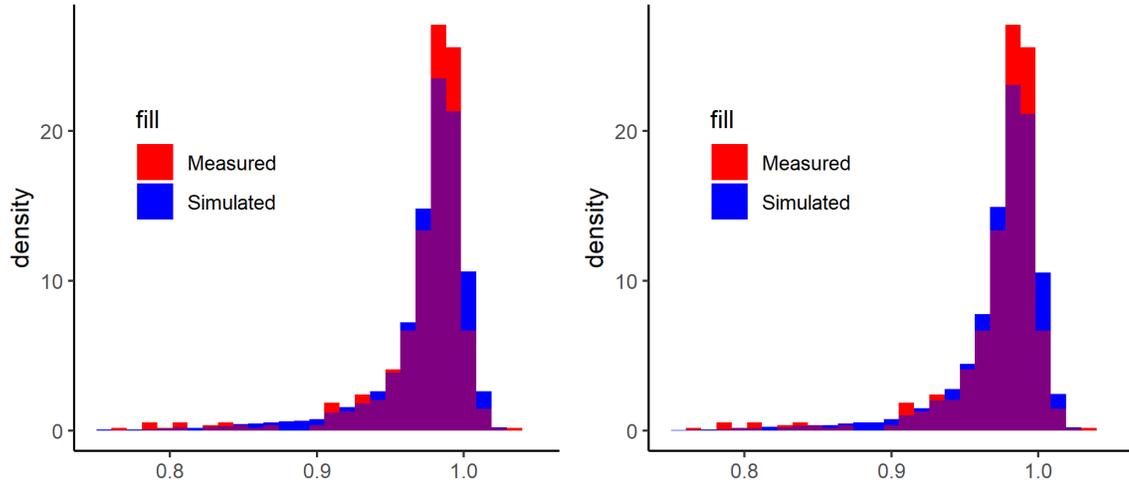


Figure 3-3 Best fits to Day 7 by simulation

The two fits of simulated to observed data, according to the Anderson-Darling test. The left and right images correspond to the first and second rows, respectively, of Table 3-4.

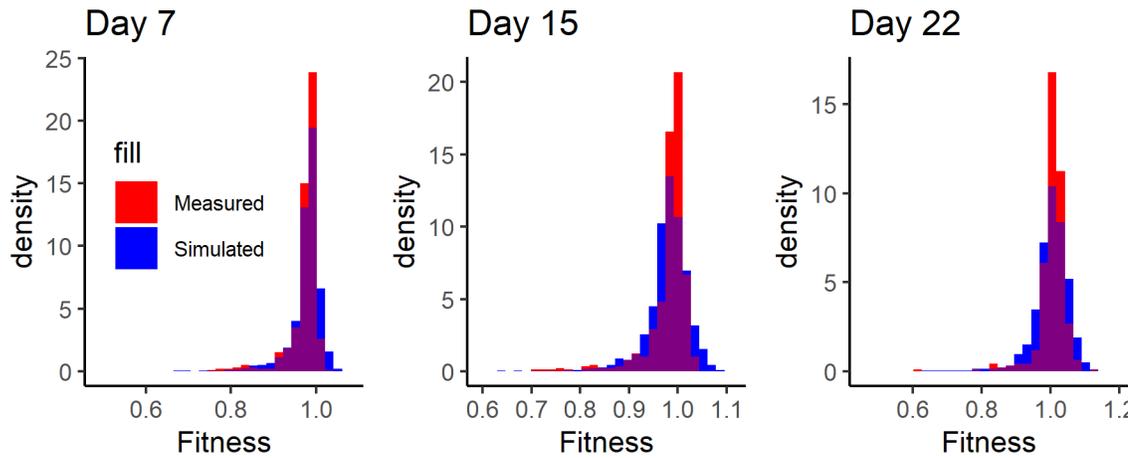


Figure 3-4 Best fit over time with beneficial mutations

In a mixture with the best-fitting deleterious DFE at Day 7 (the first row in Table 3-4), the parameters $U_{benef} = 0.002$, $\bar{s} = 0.02$, shape = 8 provided the best fit over time.

Table 3-4 Simulations of Day 7 data

All parameter combinations yielding p -values > 0.05 for the fit of the simulated to the experimental data are displayed. In total, 3644 combinations of parameters were simulated, with values spanning the following intervals: U_{del}^1 : [0.003, 0.1], \bar{s}_1 : [0.001, 0.009], $shape_1$: [0.25, 20]; U_{del}^2 : [0.003 to 0.1], \bar{s}_2 : [0.01 to 0.09], $shape_2$: [0.1, 16]. The majority of p -values ($>90\%$) were $<1e-04$.

p -value	U_{del}^1	\bar{s}_1	$shape_1$	U_{del}^2	\bar{s}_2	$shape_2$
0.092	0.05	0.005	20	0.03	0.07	0.4
0.071	0.05	0.005	12	0.04	0.05	0.25
0.069	0.05	0.005	12	0.04	0.07	0.25
0.066	0.05	0.005	20	0.03	0.09	0.4
0.060	0.05	0.005	16	0.03	0.05	0.4
0.059	0.05	0.005	16	0.03	0.07	0.4
0.056	0.05	0.007	16	0.04	0.09	0.1
0.053	0.05	0.005	16	0.03	0.09	0.4
0.052	0.05	0.005	16	0.04	0.07	0.25

3.3.5 Extending simulations to beneficial mutations

Using the best-fitting deleterious DFE as a starting point, a further 256 simulations were carried out in order to add beneficial mutations to the model, as a third gamma distribution. The best fitting distribution is shown in Figure 3-4. However, the p -value at Day 22 was quite low ($<1e-05$), so this result should not be given too much weight at present.

3.4 Discussion

This experiment clearly demonstrates that mutational load develops over time, consistent with classical population genetics theory. From Day 1 to Day 7, in the span of ~ 40 generations, the load increases from 0.3% to 2.7% as the mean fitness declines, with a long tail of less-fit individuals being added to the population. This process had not heretofore been observed in the laboratory.

It is interesting to compare these results to those obtained in the work described in Chapter 2, in which a load of 1.7% was observed through similar methodology, for similar (though not identical) haploid *mmr* yeast strains. Here we observe a higher, on Day 7 (\sim generation 50), load of 2.7%, and a lower load of 0.3% on Day 1 (\sim generation 10). This is broadly consistent with what we might assume about the number of generations experienced by the strains in Chapter 2 before plating for isolation of clones: an initial

growth in YPD after strain construction, and then two days of growth before plating, for a total of perhaps 30-35 generations.

The contributions of lethal mutations to load were not accounted for in this experiment. If we apply the lethal mutation rate for *mmr* haploid yeast as estimated in Chapter 2 (~0.004) to the Day 7 load calculated here, we obtain a lower limit of ~0.03 for the deleterious mutation rate. This substantial fitness cost may explain why mismatch-repair-deficient mutators have not been observed in natural populations of *S. cerevisiae*. In contrast, such mutators have been isolated for the relatively closely related *Candida glabrata* (Healey et al 2016), a topic explored further in the Conclusion to this thesis.

Simulations suggest that the total deleterious mutation rate is likely higher still: ~0.08, with an average selection coefficient of ~0.03. This high deleterious mutation rate is not implausible. Lang et al (2013) observed 0.85 indels per individual per generation for *mmr* haploids, of which 14% fell within coding regions. Since indels tend to cause loss-of-function mutations, a high deleterious mutation rate, even approaching 0.1, is well within the realm of possibility.

In directly comparing the loads observed in Chapter 2 with the loads observed in this experiment, one caveat is that it is possible that the pre-adaptation carried out has changed the genetics such that the Chapter 2 results are not directly comparable; but that does not seem very likely.

After Day 7, the mean fitness increases as adaptation becomes evident. (This occurred despite the pre-adaptation). Simulation suggests a beneficial mutation rate of about 0.002 (Figure 3-4), but with effect size falling off rapidly. For instance, the beneficial mutation rate for $s > 0.05$ would be about $2e-06$, within the range of a recently published estimate (Levy et al 2015), which is somewhat surprising because that experiment used wild-type, not mutator, yeast. While there may be important differences in environment between the two experiments, more analysis of the data beyond Day 7 from this experiment is needed before any firm conclusions can be drawn.

3.4.1 Summary and next steps

We have observed the development of mutational load, and subsequent adaptation, in an evolving population. We have quantified the load and drawn inferences about the rate, mean effect size, and shape of the distribution of new deleterious mutations.

Further work will include (1) applying the analytical method of Gerrish and Hengartner (2017) to the time-series fitness data in order to make inferences about the moments of the DFE, (2) extending the simulation work to further search the parameter space, with more emphasis on beneficial mutations, and (3) sequencing selected clones to identify potentially causative deleterious and beneficial mutations.

3.5 Supplemental Figures

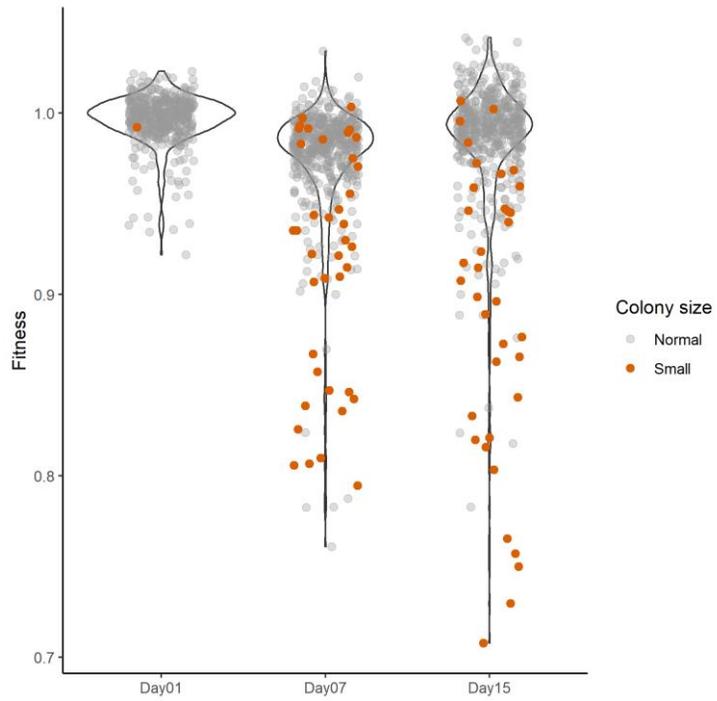


Figure 3-5 The association between colony size and fitness

Observed colony size at picking is associated with reduced fitness. Average difference in mean between normal and small size: Day 7, 0.06 ± 0.02 ($p < 1e-06$); Day 15, 0.10 ± 0.03 ($p < 1e-07$).

CONCLUSIONS AND NEXT STEPS

4.1 Conclusions

In Chapter 1, we demonstrated that recombination impedes the hitchhiking of a mutator allele, an important principle connected to the evolution of mutation rates. In Chapter 2, a substantial fitness difference (~2%) was observed for mismatch-repair-deficient (*mmr*) yeast, compared to the wild type. By measuring the fitnesses of individual clones, this bulk (*i.e.* population-to-population) difference was validated and confirmed by the within-population distribution of fitnesses. Near-direct estimation of the lethal mutation rate provided further confirmation that the observed fitness difference was due to mutational load. These results provide insight into the dynamics of evolution experiments conducted with *mmr* yeast strains; suggest a rationale for the absence, compared to bacteria, of *mmr* isolates from natural yeast populations; and provide the first, to my knowledge, unambiguous snapshot of mutational load in an evolving population.

In Chapter 3, the methodology of Chapter 2 was expanded to provide time-series data. This allowed confirmation of the high deleterious mutation rate estimated in Chapter 2, by way of estimate of an even higher deleterious mutation rate (0.03-0.08); permitted estimation of the average selection coefficient of new deleterious mutations; and provided information about the shape of the distribution of fitness effects of new mutations. Consistent with the hypothesis that *mmr* yeast experience very high indel rates, the DFE as estimated includes a substantial proportion of deleterious mutations of large effect.

These results deepen our understanding of deleterious mutation rates, mutational load, and by extension the evolution of mutation rates. While the work summarized in Chapter 3 is not yet complete and remains ongoing, in the following subsection, I detail a new strand of work that has arisen directly from what was learned in Chapters 2 and 3. In this future work, I hope to demonstrate that mutational load in the absence of mismatch repair can be partly predicted based on the composition of an organism's genome.

4.2 Predicting mutational load in a non-model organism

As mentioned in Chapter 2, homopolymeric runs (HPRs, sometimes called mononucleotide repeats in the literature) are quite mutagenic, and in particular prone to indels. For example, in *S. cerevisiae*, Tran et al (1997) reported that 1 bp deletions in a run of 7 adenines take place at a rate of $\sim 6 \times 10^{-10}$ per base, between one and two orders of magnitude higher than a recent estimate of the average per-base indel rate across the genome (Sharp et al 2018). G/C HPRs are far more mutagenic than A/T HPRs: for example, for runs of 10 identical bases, G/C HPRs appear to be about two orders of magnitude more mutagenic than A/T HPRs (Gragg et al 2002).

Mismatch repair deficiency further elevates indel rates at HPRs. In the absence of mismatch repair, the G/C indel rate is increased by about two orders of magnitude, and the A/T indel rate by about three orders of magnitude (Gragg et al 2002), for moderately long HPRs. Thus, when mismatch repair is absent, the indel rate for a base pair within an

HPR is as high as five orders of magnitude higher than for a typical base pair in the presence of functional mismatch repair.

Even at HPR lengths as short as 4, indel rates are greatly elevated compared to non-HPR sequences, and are substantially increased by the absence of a functioning mismatch repair system. In summary, the general pattern, observed in both yeast and bacteria, is that (1) HPRs are very mutagenic with respect to indels, (2) G/C HPRs are more mutagenic than A/T HPRs, and (3) the absence of mismatch repair elevates indel rates for both types dramatically, but proportionally more so for A/T HPRs (Schaaper and Dunn 1991; Tran et al. 1997; Gragg et al. 2002; Lee et al. 2012; Lang et al. 2013).

Because frameshift mutations are so likely to be damaging, it would be surprising if there were no evidence of avoidance of HPRs in genomes by codon choice. For example, consider the amino acid sequence Arginine-Glycine. The redundancy of the genetic code leads to 24 possible corresponding DNA sequences. Comparing the synonymous sequences CGA GGT and CGG GGG, the latter, having a run of 5 Gs, is substantially more mutagenic, and we might expect that such sequences would be avoided.

Gu et al (2010) performed an investigation of the genomes of over 130 species in order to distinguish such “codon pair bias” from the overall codon usage bias. Using the above example, this means that we might expect to observe the particular codon pair CGG GGG less often than expected by chance, given the overall codon usage frequencies for arginine and glycine. Indeed, the authors find strong evidence of general avoidance of runs of 5 or 6 G/Cs across 130 species, and a positive correlation between genomic GC content and avoidance of HPR-causing codon pairs, the latter finding attributed to the higher mutagenicity of G/C HPRs compared to A/T HPRs, and to the fact that as GC content increases G/C HPRs become more likely by chance.

Building on the conclusions of that study, a simple prediction is that the higher the GC content of an organism, the fewer overall HPRs (per coding base) one would expect. The reasoning is as follows: G/C HPRs are far more mutagenic than A/T HPRs, and thus more selected against, than A/T HPRs. As total GC content increases from 50%, the expected frequency of A/T HPRs (even absent any selection against them), drops, while the expected frequency of G/C HPRs rises—but we expect avoidance, by codon choice, of such sequences, as illustrated by the aforementioned Gu et al (2010) study.

To test this prediction, I downloaded the genomes for >25 prokaryotes and wrote a script to catalog HPRs in coding bases. As Figure 4-1 below shows, both within and across large taxonomic groupings of bacteria, this prediction holds.

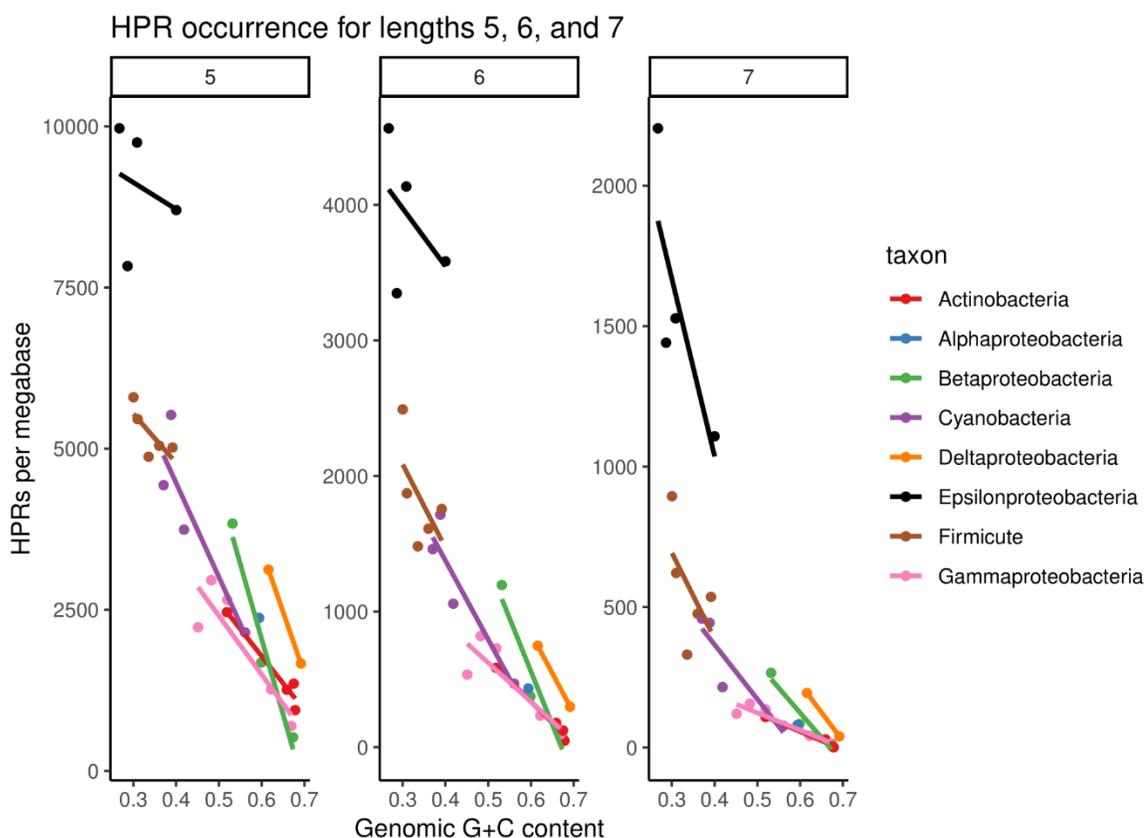


Figure 4-1 GC content and HPR occurrence for prokaryotes

The occurrence of homopolymers of length 5, 6, and 7 (both A/T and G/C) for 27 prokaryotes across various taxonomic groupings. To compute the data, annotated genomes were downloaded from public databases and HPRs counted (in coding sequences only) by a Python script. Lines are linear regression fits within taxon.

A general prediction that follows from the fact that HPRs are extremely mutagenic in the absence of mismatch repair is that the higher the HPR occurrence, the higher the mutational load incurred when mismatch repair is inactivated. In the experiments described in Chapters 2 and 3, it was established that the deleterious mutation rate for mismatch-repair-deficient *S. cerevisiae* is quite substantial—at least 0.03 and perhaps as high as 0.08, as suggested by simulation fits to time-series data. Although some small-scale mutation-rate variation for *S. cerevisiae* has been observed (Gou et al 2019), no *mmr* isolates have ever been found from natural settings. This is in contrast to the situation in *E. coli*, for which *mmr* mutators have been found repeatedly in natural isolates (reviewed in Raynes and Sniegowski 2014, and see further citations in Chapter 2), some direct evidence that the fitness cost of *mmr* is not large exists (Shaver et al 2002; Boe et al 2000; de Visser and Rozen 2006), and spontaneous *mmr* mutants have fixed in a long-term experiment (Sniegowski et al 1997). *S. cerevisiae* and *E. coli* have similar genome sizes and

similar numbers of genes, though *E. coli* has a higher GC content. In view of these facts, it is not surprising, though it is striking, that *E. coli* has a substantially lower HPR occurrence than *S. cerevisiae*, as shown in Figure 4-2A.

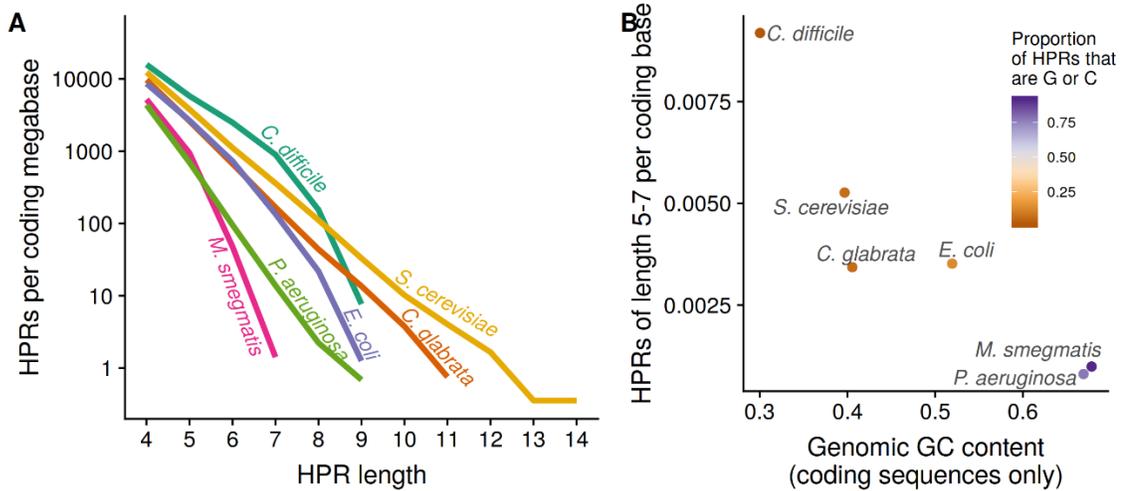


Figure 4-2 HPR occurrence vs HPR length for selected prokaryotes and yeasts

(A) HPR occurrence vs HPR length for selected organisms. Of particular interest is the difference between *S. cerevisiae* and *E. coli*. Note that *C. glabrata*, a pathogenic yeast closely related to *S. cerevisiae*, is very similar to *E. coli* in its HPR distribution. Of tangential interest: *M. smegmatis* lacks mismatch repair altogether, and appears to compensate in part by extreme avoidance of HPRs. (B) HPRs of length 5-7 vs GC content for the same organisms. *S. cerevisiae* and *C. glabrata* have similar GC content, but substantially different HPR rates.

The human pathogen *Candida glabrata* is relatively closely related (Figure 4-3) to *S. cerevisiae*. Quite interestingly, although it has similar GC content to *S. cerevisiae*, it looks much more like *E. coli* in its HPR profile (Figure 4-2A and 4-2B). A prediction, then, is that mismatch-repair-deficient *C. glabrata* would experience a lower load than mismatch-repair-deficient *S. cerevisiae*—perhaps an *E. coli*-like amount of load. Strengthening this prediction is that, much like *E. coli* and in contrast to *S. cerevisiae*, mismatch-repair-deficient isolates of *C. glabrata* have been found in clinical settings (Healey et al 2016).

This is a testable prediction, and we are currently engaged in carrying out the necessary genetic manipulations in *C. glabrata*. I plan to test this prediction using the methods developed in Chapters 2 and 3; if the prediction is confirmed, this will demonstrate that the amount of extra load experienced by a mismatch-repair mutator can be predicted by examining features of the genome.

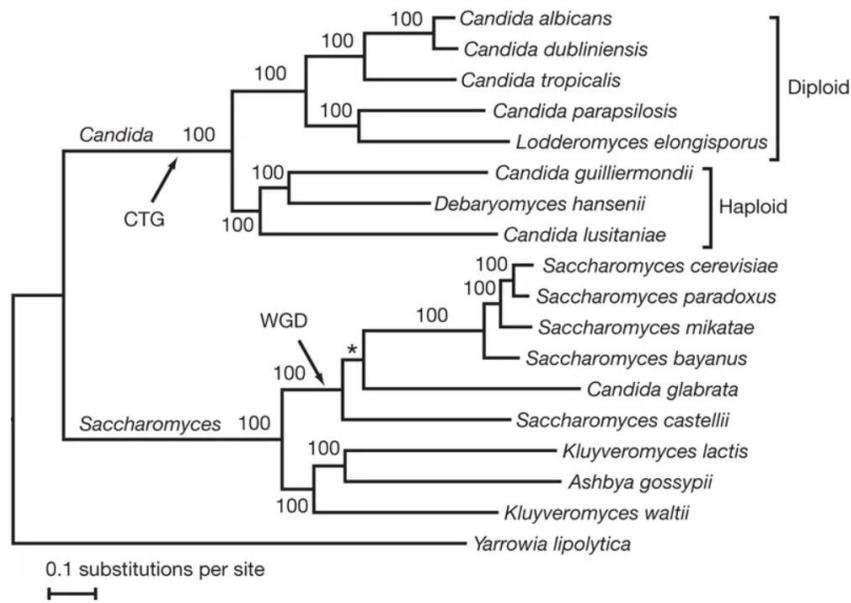


Figure 4-3 *S. cerevisiae* and *C. glabrata* are related

Yeast phylogeny, reproduced from Butler et al (2009), illustrating the relatively close relationship between *S. cerevisiae* and *C. glabrata*. Reproduced with permission from the author.

4.3 Summary

In any organism, lineages with an elevated mutation rate are generated routinely by new mutation. A modifier that elevates the mutation rate will cause increased load in its lineage. The resulting dynamics, taking beneficial mutations into account, are complex: even if mutators cannot stably invade, when back-mutation from mutator to nonmutator is accounted for, mutators can be important to the population's evolution in the long run (Desai and Fisher 2011). For a given type of mutator, such as mismatch repair deficiency, the higher the deleterious mutation rate, and the greater the portion of the deleterious DFE that is concentrated in mutations of large effect (so that mutation-selection balance is approached more quickly), the less likely such mutators are to play an important role in evolution. Both evolution experiments and observational studies suggest that mismatch-repair mutators are not likely to be important components of *S. cerevisiae* populations but may well be important for *E. coli* evolution. By grounding such results in genomic features of the organisms in question, and by demonstrating that the cost of mismatch repair deficiency can be predicted in a little-studied organism, I hope to deepen our understanding of these evolutionary processes.

REFERENCES

- Bateman, A. J. "The Viability of Near-Normal Irradiated Chromosomes." *International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine* 1, no. 2 (January 1, 1959): 170–80. <https://doi.org/10.1080/09553005914550241>.
- Beaumont, Mark A. "Approximate Bayesian Computation in Evolution and Ecology." *Annual Review of Ecology, Evolution, and Systematics* 41, no. 1 (2010): 379–406. <https://doi.org/10.1146/annurev-ecolsys-102209-144621>.
- Becks, L., and A. F. Agrawal. "The Effect of Sex on the Mean and Variance of Fitness in Facultatively Sexual Rotifers." *Journal of Evolutionary Biology* 24, no. 3 (March 1, 2011): 656–64. <https://doi.org/10.1111/j.1420-9101.2010.02199.x>.
- Butler, Geraldine, Matthew D. Rasmussen, Michael F. Lin, Manuel A. S. Santos, Sharadha Sakthikumar, Carol A. Munro, Esther Rheinbay, et al. "Evolution of Pathogenicity and Sexual Reproduction in Eight *Candida* Genomes." *Nature* 459, no. 7247 (June 2009): 657–62. <https://doi.org/10.1038/nature08064>.
- Chao, Lin, and Edward C. Cox. "Competition between High and Low Mutating Strains of *Escherichia Coli*." *Evolution* 37, no. 1 (January 1983): 125. <https://doi.org/10.2307/2408181>.
- Charlesworth, Brian. "Causes of Natural Variation in Fitness: Evidence from Studies of *Drosophila* Populations." *Proceedings of the National Academy of Sciences* 112, no. 6 (February 10, 2015): 1662–69. <https://doi.org/10.1073/pnas.1423275112>.
- Chiocchetti, Andreas, Jia Zhou, Huashun Zhu, Thomas Karl, Olaf Haubenreisser, Mark Rinnerthaler, Gino Heeren, et al. "Ribosomal Proteins Rpl10 and Rps6 Are Potent Regulators of Yeast Replicative Life Span." *Experimental Gerontology* 42, no. 4 (April 1, 2007): 275–86. <https://doi.org/10.1016/j.exger.2006.11.002>.
- Cox, Edward C., and Thomas C. Gibson. "Selection for High Mutation Rates in Chemostats." *Genetics* 77, no. 2 (1974): 169–184.
- Crow, J. F. "Genetic Loads and the Cost of Natural Selection." In *Mathematical Topics in Population Genetics*, edited by Ken-ichi Kojima, 128–77. Biomathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1970. https://doi.org/10.1007/978-3-642-46244-3_5.
- Crow, J. F., and M. Kimura. *An Introduction to Population Genetics Theory*. New York: Harper & Row, 1970. <https://www.cabdirect.org/cabdirect/abstract/19710105376>.
- Danforth, C. H. "Eugenics, Genetics and the Family: Scientific Papers of the Second International Congress of Eugenics ; Held at American Museum of Natural History, New York. September 22-28, 1921." Williams & Wilkins Company, 1923.
- Darwin, Charles. *On the Origin of Species*. London: John Murray, 1859.

- Denamur, Erick, Stéphane Bonacorsi, Antoine Giraud, Patrick Duriez, Farida Hilali, Christine Amorin, Edouard Bingen, et al. "High Frequency of Mutator Strains among Human Uropathogenic Escherichia Coli Isolates." *Journal of Bacteriology* 184, no. 2 (January 15, 2002): 605–9. <https://doi.org/10.1128/JB.184.2.605-609.2002>.
- Denamur, Erick, and Ivan Matic. "Evolution of Mutation Rates in Bacteria." *Molecular Microbiology* 60, no. 4 (May 1, 2006): 820–27. <https://doi.org/10.1111/j.1365-2958.2006.05150.x>.
- Desai, Michael M., and Daniel S. Fisher. "The Balance between Mutators and Nonmutators in Asexual Populations." *Genetics* 188, no. 4 (August 1, 2011): 997–1014. <https://doi.org/10.1534/genetics.111.128116>.
- Deschaine, Bernadette M., Angela R. Heysel, B. Adam Lenhart, and Helen A. Murphy. "Biofilm Formation and Toxin Production Provide a Fitness Advantage in Mixed Colonies of Environmental Yeast Isolates." *Ecology and Evolution* 8, no. 11 (2018): 5541–50. <https://doi.org/10.1002/ece3.4082>.
- Dobzhansky, Th, and Sewall Wright. "Genetics of Natural Populations. V. Relations between Mutation Rate and Accumulation of Lethals in Populations of Drosophila Pseudoobscura." *Genetics* 26, no. 1 (1941): 23.
- Drake, John W. "A Constant Rate of Spontaneous Mutation in DNA-Based Microbes." *Proceedings of the National Academy of Sciences* 88, no. 16 (1991): 7160–7164.
- Fisher, RA. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930.
- Galeota-Sprung, Benjamin, Breanna Guindon, and Paul Sniegowski. "The Fitness Cost of Mismatch Repair Mutators in Saccharomyces Cerevisiae : Partitioning the Mutational Load." *Heredity*, September 12, 2019, 1–12. <https://doi.org/10.1038/s41437-019-0267-2>.
- Galeota-Sprung, Benjamin, Paul Sniegowski, and Warren Ewens. "Mutational Load and the Functional Fraction of the Human Genome." *BioRxiv*, September 30, 2019, 785865. <https://doi.org/10.1101/785865>.
- Gallet, R., T. F. Cooper, S. F. Elena, and T. Lenormand. "Measuring Selection Coefficients below 10⁻³: Method, Questions, and Prospects." *Genetics* 190, no. 1 (January 1, 2012): 175–86. <https://doi.org/10.1534/genetics.111.133454>.
- Gammie, Alison E., Naz Erdeniz, Julia Beaver, Barbara Devlin, Afshan Nanji, and Mark D. Rose. "Functional Characterization of Pathogenic Human MSH2 Missense Mutations in Saccharomyces Cerevisiae." *Genetics* 177, no. 2 (October 1, 2007): 707–21. <https://doi.org/10.1534/genetics.107.071084>.
- Gentile, C. F., S.-C. Yu, S. A. Serrano, P. J. Gerrish, and P. D. Sniegowski. "Competition between High- and Higher-Mutating Strains of Escherichia Coli." *Biology Letters* 7, no. 3 (June 23, 2011): 422–24. <https://doi.org/10.1098/rsbl.2010.1036>.

- George, Carolyn M., and Eric Alani. "Multiple Cellular Mechanisms Prevent Chromosomal Rearrangements Involving Repetitive DNA." *Critical Reviews in Biochemistry and Molecular Biology* 47, no. 3 (June 1, 2012): 297–313.
<https://doi.org/10.3109/10409238.2012.675644>.
- Gerrish, Philip. "A Simple Formula for Obtaining Markedly Improved Mutation Rate Estimates." *Genetics* 180, no. 3 (November 1, 2008): 1773–78.
<https://doi.org/10.1534/genetics.108.091777>.
- Gerrish, Philip J., and Nick Hengartner. "Inferring the Distribution of Fitness Effects (DFE) of Newly-Arising Mutations Using Samples Taken from Evolving Populations in Real Time." In *Algorithms for Computational Biology*, edited by Daniel Figueiredo, Carlos Martín-Vide, Diogo Pratas, and Miguel A. Vega-Rodríguez, 103–14. Lecture Notes in Computer Science. Springer International Publishing, 2017.
- Gerstein, Aleeza C., Hye-Jung E. Chun, Alex Grant, and Sarah P. Otto. "Genomic Convergence toward Diploidy in *Saccharomyces Cerevisiae*." *PLOS Genetics* 2, no. 9 (September 22, 2006): e145. <https://doi.org/10.1371/journal.pgen.0020145>.
- Giaever, Guri, Angela M. Chu, Li Ni, Carla Connelly, Linda Riles, Steeve Véronneau, Sally Dow, et al. "Functional Profiling of the *Saccharomyces Cerevisiae* Genome." *Nature* 418, no. 6896 (July 2002): 387–91. <https://doi.org/10.1038/nature00935>.
- Gietz, R., and Robert Schiestl. "Quick and Easy Yeast Transformation Using the LiAc/SS Carrier DNA/PEG Method." *Nature Protocols* 2 (February 1, 2007): 35–37.
<https://doi.org/10.1038/nprot.2007.14>.
- Gillespie, John H. *Population Genetics: A Concise Guide*. JHU Press, 2010.
- Giraud, Antoine, Ivan Matic, Olivier Tenaillon, Antonio Clara, Miroslav Radman, Michel Fons, and François Taddei. "Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut." *Science* 291, no. 5513 (March 30, 2001): 2606–8. <https://doi.org/10.1126/science.1056421>.
- Gogarten, J. Peter, and Jeffrey P. Townsend. "Horizontal Gene Transfer, Genome Innovation and Evolution." *Nature Reviews Microbiology* 3, no. 9 (2005): 679.
- Goldstein, Alan L., and John H. McCusker. "Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*." *Yeast* 15, no. 14 (1999): 1541–53.
[https://doi.org/10.1002/\(SICI\)1097-0061\(199910\)15:14<1541::AID-YEA476>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1097-0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K).
- Gou, Liangke, Joshua S. Bloom, and Leonid Kruglyak. "The Genetic Basis of Mutation Rate Variation in Yeast." *Genetics* 211, no. 2 (February 1, 2019): 731–40.
<https://doi.org/10.1534/genetics.118.301609>.
- Gould, Carolyn V., Paul D. Sniegowski, Mikhail Shchepetov, Joshua P. Metlay, and Jeffrey N. Weiser. "Identifying Mutator Phenotypes among Fluoroquinolone-Resistant Strains of *Streptococcus Pneumoniae* Using Fluctuation Analysis." *Antimicrobial*

- Agents and Chemotherapy* 51, no. 9 (September 1, 2007): 3225–29.
<https://doi.org/10.1128/AAC.00336-07>.
- Gragg, Hana, Brian D. Harfe, and Sue Jinks-Robertson. “Base Composition of Mononucleotide Runs Affects DNA Polymerase Slippage and Removal of Frameshift Intermediates by Mismatch Repair in *Saccharomyces Cerevisiae*.” *Molecular and Cellular Biology* 22, no. 24 (December 15, 2002): 8756–62.
<https://doi.org/10.1128/MCB.22.24.8756-8762.2002>.
- Grimberg, Brian, and Clifford Zeyl. “The Effects of Sex and Mutation Rate on Adaptation in Test Tubes and to Mouse Hosts by *Saccharomyces Cerevisiae*.” *Evolution* 59, no. 2 (2005): 431–38. <https://doi.org/10.1111/j.0014-3820.2005.tb01001.x>.
- Gross, Michael D., and Eli C. Siegel. “Incidence of Mutator Strains in *Escherichia Coli* and Coliforms in Nature.” *Mutation Research Letters* 91, no. 2 (March 1, 1981): 107–10.
[https://doi.org/10.1016/0165-7992\(81\)90081-6](https://doi.org/10.1016/0165-7992(81)90081-6).
- Gu, Tingting, Shengjun Tan, Xiaoxi Gou, Hitoshi Araki, and Dacheng Tian. “Avoidance of Long Mononucleotide Repeats in Codon Pair Usage.” *Genetics* 186, no. 3 (November 1, 2010): 1077–84. <https://doi.org/10.1534/genetics.110.121137>.
- Haigh, John. “The Accumulation of Deleterious Genes in a Population—Muller’s Ratchet.” *Theoretical Population Biology* 14, no. 2 (October 1, 1978): 251–67.
[https://doi.org/10.1016/0040-5809\(78\)90027-8](https://doi.org/10.1016/0040-5809(78)90027-8).
- Haldane, J. B. S. “The Effect of Variation of Fitness.” *The American Naturalist* 71, no. 735 (July 1, 1937): 337–49. <https://doi.org/10.1086/280722>.
- Hall, David W., and Sarah B. Joseph. “A High Frequency of Beneficial Mutations across Multiple Fitness Components in *Saccharomyces Cerevisiae*.” *Genetics* 185, no. 4 (August 1, 2010): 1397–1409. <https://doi.org/10.1534/genetics.110.118307>.
- Healey, Kelley R., Yanan Zhao, Winder B. Perez, Shawn R. Lockhart, Jack D. Sobel, Dimitrios Farmakiotis, Dimitrios P. Kontoyiannis, et al. “Prevalent Mutator Genotype Identified in Fungal Pathogen *Candida Glabrata* Promotes Multi-Drug Resistance.” *Nature Communications* 7 (March 29, 2016): 11128.
<https://doi.org/10.1038/ncomms11128>.
- Henn, Brenna M., Laura R. Botigué, Stephan Peischl, Isabelle Dupanloup, Mikhail Lipatov, Brian K. Maples, Alicia R. Martin, et al. “Distance from Sub-Saharan Africa Predicts Mutational Load in Diverse Human Genomes.” *Proceedings of the National Academy of Sciences* 113, no. 4 (January 26, 2016): E440–49.
<https://doi.org/10.1073/pnas.1510805112>.
- Higgins, Kevin, and Michael Lynch. “Metapopulation Extinction Caused by Mutation Accumulation.” *Proceedings of the National Academy of Sciences* 98, no. 5 (February 27, 2001): 2928–33. <https://doi.org/10.1073/pnas.031358898>.

- Jasmin, Jean-Nicolas, and Thomas Lenormand. "Accelerating Mutational Load Is Not Due to Synergistic Epistasis or Mutator Alleles in Mutation Accumulation Lines of Yeast." *Genetics* 202, no. 2 (February 1, 2016): 751–63. <https://doi.org/10.1534/genetics.115.182774>.
- Johnson, Toby. "The Approach to Mutation–Selection Balance in an Infinite Asexual Population, and the Evolution of Mutation Rates." *Proceedings of the Royal Society of London B: Biological Sciences* 266, no. 1436 (December 7, 1999): 2389–97. <https://doi.org/10.1098/rspb.1999.0936>.
- Jyssum, K. "Observations on Two Types of Genetic Instability in Escherichia Coli." *Acta Pathologica Microbiologica Scandinavica* 48, no. 2 (1960): 113–20. <https://doi.org/10.1111/j.1699-0463.1960.tb04747.x>.
- Keightley, P. D., and Adam Eyre-Walker. "Terumi Mukai and the Riddle of Deleterious Mutation Rates." *Genetics* 153, no. 2 (October 1, 1999): 515–23.
- Kimura, Motoo. "On the Evolutionary Adjustment of Spontaneous Mutation Rates." *Genetics Research* 9, no. 1 (February 1967): 23–34. <https://doi.org/10.1017/S0016672300010284>.
- Kimura, Motoo, Takeo Maruyama, and James F. Crow. "The Mutation Load in Small Populations." *Genetics* 48, no. 10 (October 1963): 1303–12.
- Knorre, Dmitry A., Irina A. Kulemzina, Maxim I. Sorokin, Sergey A. Kochmak, Natalya A. Bocharova, Sviatoslav S. Sokolov, and Fedor F. Severin. "Sir2-Dependent Daughter-to-Mother Transport of the Damaged Proteins in Yeast Is Required to Prevent High Stress Sensitivity of the Daughters." *Cell Cycle* 9, no. 22 (November 15, 2010): 4501–5. <https://doi.org/10.4161/cc.9.22.13683>.
- Koschwanez, John H, Kevin R Foster, and Andrew W Murray. "Improved Use of a Public Good Selects for the Evolution of Undifferentiated Multicellularity." *ELife* 2 (April 2, 2013). <https://doi.org/10.7554/eLife.00367>.
- Lang, Gregory I., and Andrew W. Murray. "Estimating the Per-Base-Pair Mutation Rate in the Yeast *Saccharomyces Cerevisiae*." *Genetics* 178, no. 1 (January 1, 2008): 67–82. <https://doi.org/10.1534/genetics.107.071506>.
- Lang, Gregory I., Lance Parsons, and Alison E. Gammie. "Mutation Rates, Spectra, and Genome-Wide Distribution of Spontaneous Mutations in Mismatch Repair Deficient Yeast." *G3: Genes, Genomes, Genetics* 3, no. 9 (September 1, 2013): 1453–65. <https://doi.org/10.1534/g3.113.006429>.
- LeClerc, J. E., B. Li, W. L. Payne, and T. A. Cebula. "High Mutation Frequencies among *Escherichia Coli* and *Salmonella* Pathogens." *Science* 274, no. 5290 (November 15, 1996): 1208–11. <https://doi.org/10.1126/science.274.5290.1208>.
- Lee, H., E. Popodi, H. Tang, and P. L. Foster. "Rate and Molecular Spectrum of Spontaneous Mutations in the Bacterium *Escherichia Coli* as Determined by Whole-

- Genome Sequencing.” *Proceedings of the National Academy of Sciences* 109, no. 41 (October 9, 2012): E2774–83. <https://doi.org/10.1073/pnas.1210309109>.
- Leigh, E. G. “The Evolution of Mutation Rates.” *Genetics* 73 (April 1973): Suppl 73:1-18.
- Leigh Jr, Egbert Giles. “Natural Selection and Mutability.” *The American Naturalist* 104, no. 937 (1970): 301–305.
- Levy, Sasha F., Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. “Quantitative Evolutionary Dynamics Using High-Resolution Lineage Tracking.” *Nature* 519, no. 7542 (March 12, 2015): 181–86. <https://doi.org/10.1038/nature14279>.
- Lewontin, Richard C. *The Genetic Basis of Evolutionary Change*. Vol. 560. Columbia University Press New York, 1974.
- Li, Yuping, Sandeep Venkataram, Atish Agarwala, Barbara Dunn, Dmitri A. Petrov, Gavin Sherlock, and Daniel S. Fisher. “Hidden Complexity of Yeast Adaptation under Simple Evolutionary Conditions.” *Current Biology* 28, no. 4 (February 19, 2018): 515-525.e6. <https://doi.org/10.1016/j.cub.2018.01.009>.
- Lynch, Michael. “The Lower Bound to the Evolution of Mutation Rates.” *Genome Biology and Evolution* 3 (January 1, 2011): 1107–18. <https://doi.org/10.1093/gbe/evr066>.
- . *The Origins of Genome Architecture*. Vol. 98. Sinauer Associates Sunderland, MA, 2007.
- Lynch, Michael, Matthew S. Ackerman, Jean-Francois Gout, Hongan Long, Way Sung, W. Kelley Thomas, and Patricia L. Foster. “Genetic Drift, Selection and the Evolution of the Mutation Rate.” *Nature Reviews Genetics* 17, no. 11 (November 2016): 704–14. <https://doi.org/10.1038/nrg.2016.104>.
- Lynch, Michael, Way Sung, Krystalynne Morris, Nicole Coffey, Christian R. Landry, Erik B. Dopman, W. Joseph Dickinson, et al. “A Genome-Wide View of the Spectrum of Spontaneous Mutations in Yeast.” *Proceedings of the National Academy of Sciences* 105, no. 27 (July 8, 2008): 9272–77. <https://doi.org/10.1073/pnas.0803466105>.
- Marsischky, G. T., N. Filosi, M. F. Kane, and R. Kolodner. “Redundancy of *Saccharomyces Cerevisiae* MSH3 and MSH6 in MSH2-Dependent Mismatch Repair.” *Genes & Development* 10, no. 4 (February 15, 1996): 407–20. <https://doi.org/10.1101/gad.10.4.407>.
- Matic, Ivan, Miroslav Radman, François Taddei, Bertrand Picard, Catherine Doit, Edouard Bingen, Erick Denamur, and Jacques Elion. “Highly Variable Mutation Rates in Commensal and Pathogenic *Escherichia Coli*.” *Science* 277, no. 5333 (1997): 1833–1834.

- Mortimer, Robert K. "Evidence for Two Types of X-Ray-Induced Lethal Damage in *Saccharomyces Cerevisiae*." *Radiation Research* 2, no. 4 (June 1, 1955): 361–68. <https://doi.org/10.2307/3570244>.
- Mukai, Terumi. "The Genetic Structure of Natural Populations of *Drosophila Melanogaster*. I. Spontaneous Mutation Rate of Polygenes Controlling Viability." *Genetics* 50, no. 1 (July 1964): 1–19.
- Muller, H. J. "Our Load of Mutations." *American Journal of Human Genetics* 2, no. 2 (June 1950): III–76.
- Newcombe, Robert G. "Interval Estimation for the Difference between Independent Proportions: Comparison of Eleven Methods." *Statistics in Medicine* 17, no. 8 (1998): 873–90. [https://doi.org/10.1002/\(SICI\)1097-0258\(19980430\)17:8<873::AID-SIM779>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0258(19980430)17:8<873::AID-SIM779>3.0.CO;2-I).
- Nishant, K. T., Wu Wei, Eugenio Mancera, Juan Lucas Argueso, Andreas Schlattl, Nicolas Delhomme, Xin Ma, et al. "The Baker's Yeast Diploid Genome Is Remarkably Stable in Vegetative Growth and Meiosis." *PLOS Genetics* 6, no. 9 (September 9, 2010): e1001109. <https://doi.org/10.1371/journal.pgen.1001109>.
- Ohta, Tomoko. "Slightly Deleterious Mutant Substitutions in Evolution." *Nature* 246, no. 5428 (November 9, 1973): 96–98. <https://doi.org/10.1038/246096a0>.
- Oliver, Antonio, Rafael Cantón, Pilar Campo, Fernando Baquero, and Jesús Blázquez. "High Frequency of Hypermutable *Pseudomonas Aeruginosa* in Cystic Fibrosis Lung Infection." *Science* 288, no. 5469 (May 19, 2000): 1251–53. <https://doi.org/10.1126/science.288.5469.1251>.
- Otto, Sarah P. "The Evolutionary Enigma of Sex." *The American Naturalist* 174, no. SI (July 2009): SI–14. <https://doi.org/10.1086/599084>.
- Peter, Jackson, Matteo De Chiara, Anne Friedrich, Jia-Xing Yue, David Pflieger, Anders Bergström, Anastasie Sigwalt, et al. "Genome Evolution across 1,011 *Saccharomyces Cerevisiae* Isolates." *Nature* 556, no. 7701 (April 2018): 339. <https://doi.org/10.1038/s41586-018-0030-5>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2019. <https://www.R-project.org>.
- Raghavan, Vandana, Duyen T. Bui, Najla Al-Sweel, Anne Friedrich, Joseph Schacherer, Charles F. Aquadro, and Eric Alani. "Incompatibilities in Mismatch Repair Genes MLH1-PMS1 Contribute to a Wide Range of Mutation Rates in Human Isolates of Baker's Yeast." *Genetics* 210, no. 4 (December 1, 2018): 1253–66. <https://doi.org/10.1534/genetics.118.301550>.
- Raynes, Y., and P. D. Sniegowski. "Experimental Evolution and the Dynamics of Genomic Mutation Rate Modifiers." *Heredity* 113, no. 5 (2014): 375–380.

- Raynes, Yevgeniy, Matthew R. Gazzara, and Paul D. Sniegowski. "Contrasting Dynamics of a Mutator Allele in Asexual Populations of Differing Size." *Evolution* 66, no. 7 (July 1, 2012): 2329–34. <https://doi.org/10.1111/j.1558-5646.2011.01577.x>.
- . "Mutator Dynamics in Sexual and Asexual Experimental Populations of Yeast." *BMC Evolutionary Biology* 11, no. 1 (2011): 158.
- Raynes, Yevgeniy, C. Scott Wylie, Paul D. Sniegowski, and Daniel M. Weinreich. "Sign of Selection on Mutation Rate Modifiers Depends on Population Size." *Proceedings of the National Academy of Sciences* 115, no. 13 (March 27, 2018): 3422–27. <https://doi.org/10.1073/pnas.1715996115>.
- Reenan, R. A., and R. D. Kolodner. "Characterization of Insertion Mutations in the *Saccharomyces Cerevisiae* MSH1 and MSH2 Genes: Evidence for Separate Mitochondrial and Nuclear Functions." *Genetics* 132, no. 4 (December 1, 1992): 975–85.
- Rice, William R. "Experimental Tests of the Adaptive Significance of Sexual Recombination." *Nature Reviews. Genetics* 3, no. 4 (2002): 241.
- Richardson, Anthony R., Zhong Yu, Tanja Popovic, and Igor Stojiljkovic. "Mutator Clones of *Neisseria Meningitidis* in Epidemic Serogroup A Disease." *Proceedings of the National Academy of Sciences* 99, no. 9 (April 30, 2002): 6103–7. <https://doi.org/10.1073/pnas.092568699>.
- RStudio Team. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc., 2015. <http://www.rstudio.com/>.
- Ruderfer, Douglas M., Stephen C. Pratt, Hannah S. Seidel, and Leonid Kruglyak. "Population Genomic Analysis of Outcrossing and Recombination in Yeast." *Nature Genetics* 38, no. 9 (September 2006): 1077–81. <https://doi.org/10.1038/ng1859>.
- Schaaper, R. M., and R. L. Dunn. "Spontaneous Mutation in the *Escherichia Coli* LacI Gene." *Genetics* 129, no. 2 (October 1, 1991): 317–26.
- Serero, Alexandre, Claire Jubin, Sophie Loeillet, Patricia Legoix-Né, and Alain G. Nicolas. "Mutational Landscape of Yeast Mutator Strains." *Proceedings of the National Academy of Sciences* 111, no. 5 (February 4, 2014): 1897–1902. <https://doi.org/10.1073/pnas.1314423111>.
- Sharp, Nathaniel P., Linnea Sandell, Christopher G. James, and Sarah P. Otto. "The Genome-Wide Rate and Spectrum of Spontaneous Mutations Differ between Haploid and Diploid Yeast." *Proceedings of the National Academy of Sciences* 115, no. 22 (May 29, 2018): E5046–55. <https://doi.org/10.1073/pnas.1801040115>.
- Shaver, Aaron C., Peter G. Dombrowski, Joseph Y. Sweeney, Tania Treis, Renata M. Zappala, and Paul D. Sniegowski. "Fitness Evolution and the Rise of Mutator Alleles in Experimental *Escherichia Coli* Populations." *Genetics* 162, no. 2 (October 1, 2002): 557–66.

- Shcheprova, Zhanna, Sandro Baldi, Stephanie Buvelot Frei, Gaston Gonnet, and Yves Barral. "A Mechanism for Asymmetric Segregation of Age during Yeast Budding." *Nature* 454, no. 7205 (August 2008): 728–34. <https://doi.org/10.1038/nature07212>.
- Smith, Gerald R. "Conjugational Recombination in E. Coli: Myths and Mechanisms." *Cell* 64, no. 1 (1991): 19–27.
- Smith, John Maynard, and John Haigh. "The Hitch-Hiking Effect of a Favourable Gene." *Genetics Research* 23, no. 01 (February 1974): 23–35. <https://doi.org/10.1017/S0016672300014634>.
- Sniegowski, Paul D., Philip J. Gerrish, Toby Johnson, Aaron Shaver, and others. "The Evolution of Mutation Rates: Separating Causes from Consequences." *Bioessays* 22, no. 12 (2000): 1057–1066.
- Sniegowski, Paul D., Philip J. Gerrish, and Richard E. Lenski. "Evolution of High Mutation Rates in Experimental Populations of E. Coli." *Nature* 387, no. 6634 (June 12, 1997): 703–5. <https://doi.org/10.1038/42701>.
- Surtees, J. A., J. L. Argueso, and E. Alani. "Mismatch Repair Proteins: Key Regulators of Genetic Recombination." *Cytogenetic and Genome Research* 107, no. 3–4 (2004): 146–59. <https://doi.org/10.1159/000080593>.
- Thompson, Dawn A., Michael M. Desai, and Andrew W. Murray. "Ploidy Controls the Success of Mutators and Nature of Mutations during Budding Yeast Evolution." *Current Biology* 16, no. 16 (August 22, 2006): 1581–90. <https://doi.org/10.1016/j.cub.2006.06.070>.
- Tran, H. T., J. D. Keen, M. Krickler, M. A. Resnick, and D. A. Gordenin. "Hypermutability of Homonucleotide Runs in Mismatch Repair and DNA Polymerase Proofreading Yeast Mutants." *Molecular and Cellular Biology* 17, no. 5 (May 1, 1997): 2859–65. <https://doi.org/10.1128/MCB.17.5.2859>.
- Tröbner, Wolfgang, and Reinhard Piechocki. "Competition between Isogenic MutS and Mut+ Populations of Escherichia Coli K12 in Continuously Growing Cultures." *Molecular and General Genetics MGG* 198, no. 1 (December 1, 1984): 175–76. <https://doi.org/10.1007/BF00328719>.
- Trong, Hiep N'Guyen, Anne-Laure Prunier, and Roland Leclercq. "Hypermutable and Fluoroquinolone-Resistant Clinical Isolates of Staphylococcus Aureus." *Antimicrobial Agents and Chemotherapy* 49, no. 5 (May 1, 2005): 2098–2101. <https://doi.org/10.1128/AAC.49.5.2098-2101.2005>.
- Visser, J. Arjan G. M. de. "The Fate of Microbial Mutators." *Microbiology*, 148, no. 5 (2002): 1247–52. <https://doi.org/10.1099/00221287-148-5-1247>.
- Visser, J. Arjan G. M. de, and Daniel E. Rozen. "Clonal Interference and the Periodic Selection of New Beneficial Mutations in Escherichia Coli." *Genetics* 172, no. 4 (April 1, 2006): 2093–2100. <https://doi.org/10.1534/genetics.105.052373>.

- Wach, Achim, Arndt Brachat, Rainer Pöhlmann, and Peter Philippsen. "New Heterologous Modules for Classical or PCR-Based Gene Disruptions in *Saccharomyces Cerevisiae*." *Yeast* 10, no. 13 (1994): 1793–1808. <https://doi.org/10.1002/yea.320101310>.
- Wang, Ping, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. "Robust Growth of *Escherichia Coli*." *Current Biology* 20, no. 12 (June 22, 2010): 1099–1103. <https://doi.org/10.1016/j.cub.2010.04.045>.
- Wickham, Hadley. *Ggplot2: Elegant Graphics for Data Analysis*. Springer, 2016.
- Wielgoss, Sébastien, Jeffrey E. Barrick, Olivier Tenaillon, Michael J. Wisner, W. James Dittmar, Stéphane Cruveiller, Béatrice Chane-Woon-Ming, Claudine Médigue, Richard E. Lenski, and Dominique Schneider. "Mutation Rate Dynamics in a Bacterial Population Reflect Tension between Adaptation and Genetic Load." *Proceedings of the National Academy of Sciences* 110, no. 1 (2013): 222–227.
- Winkler, James, and Katy C. Kao. "Harnessing Recombination to Speed Adaptive Evolution in *Escherichia Coli*." *Metabolic Engineering* 14, no. 5 (September 2012): 487–95. <https://doi.org/10.1016/j.ymben.2012.07.004>.
- Wloch, Dominika M., Krzysztof Szafraniec, Rhona H. Borts, and Ryszard Korona. "Direct Estimate of the Mutation Rate and the Distribution of Fitness Effects in the Yeast *Saccharomyces Cerevisiae*." *Genetics* 159, no. 2 (October 1, 2001): 441–52.
- Zeyl, Clifford, and J. Arjan G. M. DeVisser. "Estimates of the Rate and Distribution of Fitness Effects of Spontaneous Mutation in *Saccharomyces Cerevisiae*." *Genetics* 157, no. 1 (January 1, 2001): 53–61.
- Zhu, Y. O., M. L. Siegal, D. W. Hall, and D. A. Petrov. "Precise Estimates of Mutation Rate and Spectrum in Yeast." *Proceedings of the National Academy of Sciences* 111, no. 22 (June 3, 2014): E2310–18. <https://doi.org/10.1073/pnas.1323011111>.